MOLECULAR AND SEROLOGICAL CHARACTERIZATION OF WATERMELON LEAF MOTTLE VIRUS (WLMV)

By
PATRICIA DE SA GUIMARAES

A DISSERTATION PRESENTED TO THE GRADUATE SCHOOL
OF THE UNIVERSITY OF FLORIDA IN PARTIAL FULFILLMENT
OF THE REQUIREMENTS FOR THE DEGREE OF
DOCTOR OF PHILOSOPHY

UNIVERSITY OF FLORIDA

1999

ACKNOWLEDGEMENTS

I would like to thank the 'Conselho Nacional de Pesquisa e Pos Graduacao-CNPq' the National Council for Research and Post-Graduate Studies of Brazil, for sponsoring me. Their support has enabled me to pursue my interest in Plant Pathology.

I would like to express sincere gratitude to Drs. Ernest Hiebert, Dan E. Purcifull, Francis W. Zettler and William B. Gurley, members of my supervisory committee for guidance, support and encouragement throughout this study. I would like to thank Dr. George N. Agrios for his support of my research.

I would like to express sincere thanks to Dr. William Farmery from the Molecular Genetics Core for helpful discussions on PCR and protein expression, and Dr. Linda Green and the Hybridoma Core Laboratory for help with production of ascitic fluid. Special thanks are due to Dr. Savita Shanker and the DNA Sequencing Core Laboratory for sequencing expertise and advice

I am deeply grateful for the love and loyal support of my children, Bruno, Nara and Marcos, who had to contend with a student-mother through their teenage years.

I am deeply grateful to Kelly Snow for his constant and unfaltering support, and encouragement at all times.

I thank Mark Elliott, Margaret Smithers-Kopperl, Ahmed Abouzid, Kristin Beckham, Eugene Crawford, Xandrika Ramadugu and Manjunath Keremane for their helpful discussions and friendship.

I would like to thank Dr. Rachel Shireman for the learning opportunity I had as a member of the Graduate Student Advisory Committee, and for her genuine interest and support of me, and of all the graduate students, in her role as IFAS Associate Dean of Graduate Programs.

TABLE OF CONTENTS

page

ACKNOWLEDGEMENTS....ii

ABSTRACT....vi

CHAPTERS

1. INTRODUCTION....1

General Characteristics of the *Potyvirus* Genus....1

Research Objectives....17

2. MOLECULAR CHARACTERIZATION OF THE COAT PROTEIN OF WATERMELON LEAF MOTTLE VIRUS (WLMV)....20

Introduction....20

Materials and Methods....21

Results....27

Discussion....29

3. CLONING AND COMPLETE NUCLEOTIDE SEQUENCING OF WATERMELON LEAF MOTTLE VIRUS....43

Introduction....43

Materials and Methods....45

Results....56

Discussion....63

4. SEROLOGICAL CHARACTERIZATION OF WATERMELON LEAF MOTTLE VIRUS....84

Introduction....84

Materials and Methods....86

Results....97

Discussion....103

5. SUMMARY AND CONCLUSIONS....120
LITERATURE CITED....126
BIOGRAPHICAL SKETCH....136

Abstract of Dissertation Presented to the Graduate School of the University of Florida in Partial Fulfillment of the Requirements for the Degree of Doctor of Philosophy

MOLECULAR AND SEROLOGICAL CHARACTERIZATION OF WATERMELON LEAF MOTTLE VIRUS (WLMV)

By

Patricia de Sa Guimaraes

May 1999

Chairman: Ernest Hiebert
Cochairman: Dan E. Purcifull
Major Department: Plant Pathology

Biological, serological and molecular studies were conducted with watermelon leaf mottle virus (WLMV), a newly identified potyvirus infecting watermelon in Florida. The entire genome of WLMV was cloned and sequenced. The genome consists of a single open reading reading frame (ORF) of 9,657 nucleotides (nt) preceded by a 5' untranslated region of 94 nt and terminated by a 3' untranslated region (3'UTR) of 243 nt.

The taxonomic status of WLMV was determined by comparisons of the sequence of the cloned viral genome with potyvirus sequences obtained from GenBank. Percentage of sequence identity for coat protein amino acids of WLMV and of 17 other potyviruses ranged from 54.8 to 65.2%. The nucleotide sequence identities ranged from 55.6 to 63.5% for the coat protein, and from 37.2 to 48.3% for the 3'UTR. The percentage of

sequence identity between the polyproteins of WLMV and of nine other potyviruses ranged from 44.7% to 48.4 %.

Phylogenetic analysis of the coat protein region and the 3'UTR, using neighbor-joining and maximum parsimony methods, consistently indicated that WLMV did not cluster with other potyviruses in a clade with high bootstrap support.

The coat protein gene of WLMV was PCR amplified and expressed in *Escherichia coli*. The expressed coat protein was used for production of a polyclonal antiserum in rabbit and ascitic fluid in mice. The antiserum produced in rabbit was a suitable serological probe for detection of WLMV in SDS immunodiffusion, indirect ELISA, and Western blots. The ascitic fluid was useful for detection of WLMV in indirect ELISA and in Western blots. PTY-1 and ADMIX monoclonal antibodies did not reliably detect WLMV in indirect ELISA or Western blots.

Consistent and clear symptoms resulting from inoculation with WLMV were found in watermelon 'Crimson Sweet' and 'Charleston Gray', in 'Early Prolific Straightneck' and 'Prelude II' squash, and in *Cucumis metuliferus* PI 292109, PI 482459 and Accession number 2459. *Luffa acutangula*, *Lagenaria siceraria*, *Cucurbita okeechobeensis* var *okeechobeensis*, and some commercial varieties of cucumber and melon were either symptomless hosts, or showed mild symptoms.

Sequence comparisons and phylogenetic analysis indicated that watermelon leaf mottle virus is a distinct member of the *Potyvirus* genus.

CHAPTER 1
INTRODUCTION

General Characteristics of the *Potyvirus* Genus

Potyviridae Family

The Potyviridae family comprises three genera, *Potyvirus*, *Bymovirus* and *Rymovirus*, and unassigned viruses grouped in two possible genera, *Ipomovirus* and *Macluravirus*. The *Ipomovirus* genus includes sweet potato mild mottle virus (SPMMV) and sweet potato yellow dwarf virus, and the viruses in the genus are vectored by the whitefly *Bemisia tabaci*. The *Macluravirus* genus includes the aphid transmitted Maclura mosaic virus (MacMV) and narcissus latent virus (NLV) (Shukla *et al.*, 1994). All the viruses within the Potyviridae family have filamentous particles, and except for the genus *Bymovirus*, are monopartite. The accumulation of proteinaceous cylindrical inclusions in the cytoplasm of infected plants is a diagnostic feature for the family (Edwardson, 1992).

The type virus in the genus *Rymovirus* is ryegrass mosaic virus (RGMV), and the virus particles measure from 650 to 900 nm in length. They are vectored by the eriophyid mites *Abacaris hystrix* and *Aceria tulipae* and have natural and experimental host ranges which may include from ten to over a hundred species of plants that are mostly limited to the family Gramineae. Viruses in the genus *Bymovirus* are bipartite and have particles of two modal lengths, varying from 250 to 300 nm and from 500 to 600 nm. They are

transmitted by the fungus *Polymixa graminis* (Plasmodiophoraceae) and have limited host ranges within the family Gramineae. The type member of the genus is barley yellow mosaic virus (BaYMV) (Murphy *et al.*, 1995).

The word "potyvirus" is an acronym for potato virus Y (PVY), the type member of the genus *Potyvirus* that includes aphid-transmitted viruses with flexuous, filamentous particles with helical symmetry. Virus particles measure from 680 to 900 nm in length and from 11 nm to 13 nm in width. A few potyviruses have broad host ranges infecting plant species in many plant families, but most have narrow host ranges. The Sixth Report of the International Committee on Taxonomy of Viruses (ICTV) lists 75 definitive and 93 tentative species in the genus (Murphy *et al.*, 1995). The *Potyvirus* genus is the largest of the 47 plant virus groups recognized by the ICTV; new members are described continually and the taxonomic status of recognized species and strains continues to change as improved methods for the molecular and taxonomical characterization of viruses are used.

Organization of the *Potyvirus* Genome

The *Potyvirus* genome is a single-stranded linear molecule of positive sense RNA with approximately 10,000 nucleotides (nt). The genome has a single open reading frame (ORF) which is translated into one precursor polyprotein. A 5' untranslated region (5'UTR) precedes the ORF, and the 5' terminal nucleotide is covalently linked to a tyrosine residue of the genome-linked protein (VPg). The ORF is followed by a 3' untranslated region (3'UTR) and the poly (A) tail, a poly adenylated tract of varying length.

The precursor polyprotein undergoes cotranslational and posttranslational cleavage by three virus encoded proteases into nine structural and nonstructural proteins: P1

protease; helper component protease (HC-Pro); P3 protein; 6K1 fragment; cylindrical inclusion protein (CI); 6K2 fragment; VPg protease (VPg-Pro), the replicase and the coat protein (CP). The proteins responsible for the proteolytic cleavages are the P1 protease, the HC-Pro and the VPg-Pro (Riechman *et al.*, 1992). There is no differential regulation of the amount of protein produced during translation of the potyvirus RNA, and nonstructural proteins (NSP) may be produced in excess of the functionally needed amounts.

Some NSP accumulate in the cells forming inclusions that have been suggested to be sites of accumulation of excess or partially processed polyproteins that are inactive (Martin and Gelie, 1997; Riedel *et al.*, 1998). Inclusions are found in the cytoplasm and in the nucleus, and many are formed by the accumulation of more than one protein. Functional NSP are expected to be found in the cytoplasm in very small amounts, nonaggregated, and probably associated with membranes (Riedel *et al.*, 1998). Glycerol gradient fractions of plant tissue infected with plum pox virus (PPV), and with RNA polymerase activity, cosedimented with the endoplasmic reticulum and contained NIa, NIb and CI proteins (Martin and Gelie, 1997).

Untranslated Regions

The 3'UTR is heterogeneous in size and sequence, contains AU-rich sequences, GAGGN/N'CCUC inverted repeats, and it can be predicted to fold into stable secondary structures (Shukla *et al.*, 1994). The secondary structures formed by the 3'UTR of potyviruses have been implicated in symptom severity and in genome amplification. Changes in the length of the 3'UTR of tobacco vein mottling virus (TVMV) resulted in symptom attenuation without reducing genome amplification, and were associated with

changes in secondary structure patterns of the 3'UTR (Rodriguez-Cerezo *et al.*, 1991). Mutations in the CP coding region and in the 3'UTR of tobacco etch virus (TEV) that resulted in changes in the secondary structure, reduced genome amplification of TEV (Haldeman-Cahill *et al.*, 1998).

The poly (A) tail of a mRNA is a regulator of expression, it regulates translation by increasing translational efficiency and by maintaining mRNA stability in plant and animal cells. For its role in stimulation of translational efficiency, the poly (A) tail is functionally dependent on the 5' cap structure (Gallie, 1991). Gallie *et al.* (1995) demonstrated that in the absence of a cap, the poly (A) tail could stimulate translation of a mRNA if this messenger RNA contained the 5'UTR of TEV and concluded that the TEV 5'UTR and the poly (A) tail functioned synergistically. The length of the 3'UTR may play an important role in determining the extent to which a poly (A) tail can stimulate the translation of a mRNA (Tanguay and Gallie, 1996).

The 5'UTR has been shown to act as a translational enhancer of the viral RNA in plant cells (Carrington and Freed, 1990; Nicolaisen *et al.*, 1992). Basso *et al.* (1994) suggested that ribosomes bind to an internal ribosomal entry site (IRES) within the 5'UTR of turnip mosaic virus (TuMV) and scan the sequence for the initiator AUG. Translation of PPV was shown to be independent of secondary structures in the 5'UTR. However, the recognition of the AUG triplet was dependent on its placement in an appropriate sequence context and distance from the 5' end of the genome. Cap independent leaky scanning was proposed as the mechanism for translation initiation of PPV RNA (Simon-Buela *et al.*, 1997a).

The 5'UTR has conserved sequences known as the "potybox" and "box a" and "box b", and these sequences may be important for processes such as encapsidation, translation or replication (Shukla *et al.*, 1994). The 5'UTR seems to be devoid of stable secondary structure and the 5' terminal 31 nucleotides are conserved in all potyviruses. Through deletion and mutational study of segments of the 5'UTR of PPV, this 31 nucleotide region was proposed as a target for viral and possibly host factors involved in virus replication. These conserved nucleotides could be the target for the replication complex or could participate in the encapsidation process (Simon-Buela, *et al.*, 1997b). The potybox and the boxes a and b are within the 31 nucleotide region.

The P1 Protein

The P1 protein is the amino terminal protein of the polyprotein, and it is a serine type protease which catalyses the autoproteolytic cleavage between itself and HC-Pro. P1 is the least conserved protein among potyviruses with a sequence identity of 18 to 19%. The amino terminal region is highly divergent, whereas the carboxy terminal has highly conserved amino acids. The motif GXSG is strictly conserved in potyvirus P1 proteins and is found in cellular serine type proteases. The histidine and aspartate residues found upstream of the conserved active site serine in the GXSG motif and believed to be components of the active center, are also strictly conserved (Verchot *et al.*, 1991). The FIVRG motif is also highly conserved in potyviruses and is believed to be important for protein processing at the junction between P1 and HC-Pro (Verchot *et al.*, 1991).

The role of the P1 protein in virus infectivity is not completely understood, but some functions associated with this protein have been identified. The results of mutational and complementation studies using a TEV mutant, from which the entire P1 coding

sequence was deleted, indicate that the P1 protein stimulates genome amplification in *trans* as an accessory factor. Genome amplification was reduced, but cell to cell and long distance transport were only minimally affected (Verchot and Carrington, 1995). The P1 proteins of TuMV and of tobacco vein mottling virus (TVMV) have also been shown to be RNA binding proteins (Soumounou and Laliberte, 1994; Brantley and Hunt, 1993).

The P1 protein has been found in the cytoplasm and in the nucleus of infected cells. The P1 protein of TVMV was found in membrane-enriched fractions of extracts of infected *Nicotiana tabacum* (Rodriguez-Cerezo and Shaw, 1991). The P1 of zucchini yellow mosaic virus (ZYMV) was localized using fluorescence microscopy and antisera to the P1 protein labeled with protein A and conjugated to rhodamine. The P1 proteins of four isolates of ZYMV were detected as amorphous particulate aggregates in the cytoplasm of cells in epidermal strips from infected watermelons (Wisler, 1992). Immunogold labeling using homologous polyclonal antibodies localized the P1 protein of PVY associated to the CI and in the cytoplasm of infected *N. tabacum* leaf cells. No P1 was found elsewhere in the cells (Arbatova *et al.*, 1998). However, the P1 protein of the Rankovic strain of PPV (PPV-RA) was shown by immunogold labeling to be associated with nuclear inclusions and the dense material located in the cytoplasm of *Nicotiana clevelandii* (Martin and Gelie, 1997).

Helper Component Protease

The helper component protease (HC-Pro) is a papain-like protease with an active cysteine residue that autoproteolytically cleaves itself from the P3 protein. Proteolysis occurs between two glycine residues within the conserved YXVGG motif (Carrington and Herndon, 1992; Maia *et al.*, 1996). HC-Pro is a multifunctional protein. Genetic and

biochemical evidence indicate that the amino terminal region of HC-Pro is required for aphid transmission and for efficient genome amplification. The central region is important for genome amplification and long distance movement in plants and may involve the conserved motif CCCE. The carboxy terminal domain is the protease domain (Kasschau *et al.*, 1997). The amino terminal region has a strongly conserved cysteine-rich region postulated to form a zinc finger-like motif, with structural or functional roles, and possibly associated with nucleic acid binding, RNA folding, and dimerization. The lysine residue within this region may participate in an ionic interaction required for the formation of biologically active HC-Pro dimers (Atreya *et al.*, 1992). The KITC motif within the N-terminal region is required for interaction of HC-Pro with aphid stylets, and mutations involving charge changes within or near the cysteine rich region surrounding the KITC motif resulted in loss of activity (Blanc *et al.*, 1998). Although the KITC motif is common to most potyviruses, mutations from K to R are tolerated, and the sequence KLSC is found in ZYMV and in soybean mosaic virus (SMV) (Huet *et al.*, 1994). Virus strains mutated in the PTK motif that rendered HC-Pro nonfunctional in transmission failed to bind to virions. Activity of the PTK motif was attributed to a structural interaction between HC-Pro and the virus particles, which may also occur in cells. The conserved motifs KITC and PTK must be present in the same HC-Pro molecule for biological activity, suggesting that the bridge model is the most likely explanation for the mode of action of HC-Pro (Peng *et al.*, 1998).

Homology-dependent gene silencing in plants can suppress RNA accumulation in a sequence dependent way. However, infection by a plant virus can induce posttranscriptional or trancriptional gene silencing in the absence of homology of the virus

genome to host genes. Gene silencing may function as an adaptive antiviral defense response, where virus induced silencing conditions an immune state in which subsequent infection by closely related strains is restricted. The P1/HC-Pro sequence of TEV has been proposed to function as a suppressor of gene silencing. The P1/HC-Pro sequence is also believed to stimulate or prolong amplification of viruses that are not related to TEV, such as potato virus X (PVX). PVX accumulates to higher levels in the presence of HC-Pro from a coinfecting TEV, a transgene or a recombinant strain (Anandalakshmi *et al.*, 1998; Kasschau and Carrington, 1998). The central domain of HC-Pro was found to be required for synergistic disease between TEV and PVX in experiments using several lines of transgenic *N. tabacum* expressing mutated HC-Pro (Shi *et al.*, 1997). The expression of the P1/HC-Pro sequence enhanced the accumulation and pathogenicity of tobacco mosaic virus (TMV) and cucumber mosaic virus (CMV) and PVX (Pruss *et al.*, 1997).

The HC-Pro accumulates and forms amorphous inclusions (AI) in the cytoplasm of cells infected with some potyviruses including papaya ringspot type W (PRSV-W), PVY, TVMV, TuMV, and PPV, but not TEV (Shukla *et al.*, 1994; Riedel *et al.*, 1998). HC-Pro has also been found in the nucleus associated with nuclear inclusions in beet mosaic virus (BtMV) and TEV (Riedel *et al.*, 1998), or as a layered inclusion in squash and watermelon infected with watermelon leaf mottle virus (WLMV) (Petersen *et al.*, 1996).

P3 Protein

Computer analysis of the deduced amino acid sequence of the TVMV P3 protein has indicated that it is probably an integral transmembrane protein, and P3 has been found predominantly in membrane enriched fractions of TVMV infected tobacco leaves (Rodriguez-Cerezo and Shaw, 1991). Computer analysis of the deduced amino acid

sequences of P3 proteins of ZYMV, PRSV-W and WLMV indicate that they are transmembrane proteins (Wisler, 1992; De Sa, unpublished). The amino terminal region of P3 proteins is more conserved, and the sequence $EPYX_7SPX_2LXAX_2NXGX_2EX_5W$ is similar to the sequence $EPFX_{11}VXAX_3NXGX_8M$ found in a corresponding position in the 32 kDa protein of cowpea mosaic virus (CPMV). This has led to the suggestion that P3 could be involved in the regulation of proteolytic processing of the polyprotein (Riechmann *et al.*, 1992).

Immunogold labeling using polyclonal antibodies prepared to the P3 has been used to localize the P3 protein of different potyviruses. The P3 of TVMV was found associated with the CI in infected *N. tabacum* (Rodriguez Cerezo *et al.*, 1993), and with an accumulation of dense material in the cytoplasm of in *N. clevelandii* infected with PPV-RA (Martin and Gelie, 1997). P3 was also found associated with the nuclear inclusions in *N. clevelandii* infected with PPV-RA (Martin and Gelie, 1997), and in *N. tabacum* infected with TEV (Langenberg and Zhang, 1997).

Cylindrical Inclusion Protein

All potyviruses induce cylindrical inclusion protein (CI), which accumulates in the cell cytoplasm as cylindrical inclusions. The CI protein, believed to be a multifunctional protein, was found to have nucleic acid-stimulated NTPase activity related to helix unwinding in PPV, and in tamarillo mosaic potyvirus (Lain *et al.*, 1991; Eagles *et al.*, 1994). The helicase and ATPase activities were also detected in virus-induced vesicles in TEV infected tissues, but these activities were not directly linked to the CI. This finding led to the suggestion that the CI protein exists in an active form associated with membranes and vesicles, and in a nonenzymatic inclusion form. (Chen *et al.*, 1995). The

motif GAVGSGKST is highly conserved in all potyviruses and is in agreement with the nucleoside triphosphate binding motif A/GXXXXGKS/T found in ATPases (Lain *et al.*, 1991). Two other motifs are conserved in the CI proteins of potyviruses: DEXH and QXXGRXXR. The DEXH motif is believed to be involved in NTPase activity and to interact with the QXXGRXXR motif, which is involved in RNA binding in helicase-like proteins. The latter is within a highly conserved region in all potyviruses (Fernandez and Garcia, 1996).

The CI protein has been shown to act transiently in aiding virus movement. An ultrastructural study of the advancing infection in pea cotyledons infected with pea seed borne mosaic virus (PSbMV) showed that the CI protein exists in transient functional states (Roberts *et al.*, 1998). Pinwheel inclusions were observed in the initial stages of infection positioned over plasmodesmata, and virus coat protein was found within the modified plasmodesmata. This was also linked with a transient reduction in callose in the vicinity of the plasmodesmata. In later stages of infection, the CI were no longer associated with the cell wall or with the coat protein, and they accumulated in the cytoplasm as pinwheel aggregates typical of all potyvirus infections (Roberts *et al.*, 1998).

6K1 and 6K2 Proteins

The two small proteins 6K1 and 6K2 have hydrophobic stretches of amino acids and may be involved in potyvirus replication (Riechman *et al.*, 1992). Infection by TEV was associated with collapse of the endoplasmic reticulum into aggregated vesicles, and viral replication complexes were associated with ER-like membranes. The 6K2 protein of TEV is an integral membrane protein with a membrane targetting motif consisting of a hydrophobic domain and flanking charged amino acids, and is found associated with

vesicular membranes derived from the endoplasmic reticulum. The 6K2 protein of TEV has been proposed to function as an anchor to secure the RNA replication complex, or parts of it, to membranous sites (Schaad *et al.*, 1997a).

Genome-Linked Protein Protease

The major viral protease is the genome-linked protein protease (VPg-Pro) consisting of a carboxy terminal protease and an amino terminal genome-linked structural protein. The two domains are separated by an internal suboptimal cleavage site, which is believed to be inefficiently processed. VPg-Pro is similar to the cysteine-type proteases structurally related to the trypsin-like serine proteases. The active site sequence GXCG is conserved in all potyviruses, and the cleavage sites have a general potyvirus-specific consensus of VXXQ/(A, S, G, V). Virus specific motifs are superimposed on this sequence and affect the rate of cleavage (Shukla *et al.*, 1995). VPg-Pro processing of the polyprotein is a regulated process involving *cis* and *trans* cleavages at sites recognized with different levels of efficiency. The expression of potyviral gene products is not differentially regulated at the levels of transcription and translation, but may be regulated by sequential proteolytic events (Riechman *et al.*, 1992). Cleavage at one site may modify the conformation of the polyprotein, changing the efficiency of cleavage at another site and establishing the order of processing events. Other viral or plant co-factors may participate in the regulation of polyprotein processing (Garcia *et al.*, 1992).

The carboxy terminal region of VPg-Pro is less conserved than the amino terminal region, and it has been suggested that the carboxy terminal region is involved in forming a substrate pocket. Based on construction of hybrid proteases, it has been suggested that the main determinants for substrate specificity lie in the carboxy terminal one-third of the

protease, and that the recognition and the catalytic sites are closely linked (Riechmann *et al.*, 1992).

The amino terminal domain of the VPg-Pro is the genome-linked protein (VPg). The VPg is attached covalently to the 5' terminus of genomic RNA through a phosphodiester linkage with the hydroxyl group of a tyrosine residue (Murphy *et al.*, 1991; Murphy *et al.*, 1996).

Various forms of the VPg are associated with the viral RNA. PPV and TVMV RNAs are found associated with the VPg (Keller *et al.*, 1998), and nuclease digestion experiments with TEV RNA have released products containing both the processed VPg and the full-length VPg-Pro. Mutations that abolish or accelerate the cleavage between VPg and the protease domain are detrimental to genome amplification. Slow processing at the internal site may serve a regulatory function by controlling the ratio between processed and full-length forms of VPg-Pro, or may result in timed activation or deactivation of its functions (Schaad *et al.*, 1996).

The interactions between a potyvirus strain and host cultivars resistant to that strain are used to study the pathogenicity determinants of potyviruses and to clarify protein functions in infection. The mechanism of resistance of *Pisum sativum* to pea seed borne mosaic virus (PSbMV) was investigated, and the VPg was identified as the determinant of pathotype-specific virulence of PSbMV in the *sbm-1/sbm-1* genotype of pea. VPg is believed to be involved in virus replication, and this may be disrupted by the lack of recognition and interaction between a host component and the VPg (Keller *et al.*, 1998). In studies using TEV and *N. tabacum*, VPg was shown to be a host genotype-specific determinant for long-distance movement. It was suggested the VPg interacts

directly or indirectly with host components to facilitate long-distance movement (Schaad *et al.*, 1997b).

RNA-Dependent RNA Polymerase

Approximately 70% of plant viruses are plus-stranded RNA viruses and the genomic RNA can be directly translated into viral proteins (Mathews, 1992; Ishihama and Barbier; 1994). The RNA-dependent RNA polymerase, the replicase, participates in viral RNA (vRNA)-directed synthesis of complementary RNA (cRNA) of negative polarity and cRNA-directed synthesis of vRNA. Isolation and characterization of replicases from infected cells have been difficult due to two factors: low extractability of active replicases from membrane complexes together with their strong association with endogenous RNA templates, and the presence of an inducible host RNA dependent RNA polymerase activity (Ishihama and Barbier, 1994). Experimental techniques for gene cloning and sequencing allow for the identification of nucleotide sequences of the RNA genome and production of viral proteins from cDNA clones. Sequence analysis of coding sequences for the replicases of various viruses indicates multiple functional domains and motifs (Ishihama and Barbier, 1994).

The sequence of the replicase is well conserved between different potyviruses and shows identities from 61 to 64% (Shukla *et al.*, 1994). Conservation is higher at the amino terminal region, and there are two conserved motifs characteristic of RNA-dependent RNA polymerases: (T/S)GXXXTXXXN(T/S) 21 to 38 aa GDD with the sequence GNNSGQPSTVVDNT (30 aa) GDD (Domier *et al.*, 1987). The GDD sequence is a hallmark of viral RNA-dependent RNA polymerases (Ishihama and Barbier, 1994). Hong and Hunt (1996) demonstrated that the NIb protein of TVMV is the replicase and that the

GDD motif is important for polymerase activity. It has also been shown by genetic complementation analysis that the NIb protein of TEV is a trans-active protein that can be supplied from outside the context of the TEV polyprotein. Protein-protein interactions with host or virus factors or protein-RNA interactions may recruit the replicase to the replication complexes. Using a yeast two-hybrid system, the protease domain of VPg-Pro of TEV was shown to be the primary determinant in an interaction with the replicase. A model for the interaction of the VPg-Pro and the replicase during initiation of synthesis of TEV vRNA was devised in which the 6K/VPg-Pro complex is anchored to membranes, and the replicase is recruited to the complex by an interaction with the protease domain of VPg-Pro. Initiation of RNA synthesis is stimulated by a priming mechanism involving the VPg domain of VPg-Pro (Li *et al.*, 1997).

The VPg-Pro and the replicase accumulate in nearly equimolar amounts, forming nuclear inclusions in plant cells infected by several potyviruses. In some viruses, such as TEV, the nuclear inclusions are detected by light microscopy with adequate staining (Christie and Edwarson, 1994). The VPg-Pro and the replicase of peanut stripe virus (PStV) accumulate in the nucleus without forming inclusions (Hajimorad *et al.*, 1996), and this may be true for other potyviruses. The nuclear inclusion a (NIa) and the nuclear inclusion b (NIb) are made up primarily of VPg-Pro and the replicase, respectively. However, Riedel *et al.* (1998) have shown that the crystalline nuclear inclusions may also contain other proteins. TEV nuclear inclusions were made up of NIa, NIb and HC-Pro, and BtMV nuclear inclusions were made up of NIa, NIb, HC-Pro and CI. This may be an indication that some incompletely processed proteins accumulate in inclusions.

Entry of proteins into the nucleus is a complex process that involves energy independent recognition and binding of a nuclear localization signal (NLS) to the nuclear pore complex. Uptake into the nucleus through the pore is an energy-dependent process that requires ATP hydrolysis (Hicks and Raikhel, 1995). The NLS are composed primarily of basic residues. The main types are: the SV40-like (PKKKRKV), consisting of a short sequence of seven or eight amino acids containing a single cluster of basic residues and frequently flanked by a prolyl or acidic amino acid; the MAT α2-like, consisting of short hydrophobic regions with one or more hydrophobic amino acids; and the bipartite NLS with a combination of two regions of basic amino acids separated by a spacer of four or more residues (SPPKAVKRPAATKKAGQAKKKKLDKEDES). The SV40-like NLSs are divided into nine typical NLS motifs, and three non-typical motifs. There are also other examples of NLS. The NLS and DNA/RNA-binding domains of the majority of nuclear proteins either overlap or flank each other, and this is thought to be important for direct localization of the proteins to the site of function (LaCasse and Lefebvre, 1995).

The VPg-Pro and the replicase of TEV contain independent nuclear localization signals. The VPg-Pro has a nuclear localization signal within the VPg domain. Constructs carrying a β - glucoronidase (GUS) reporter gene with mutations affecting the transport of GUS/VPg to the nucleus had significant effects on genome amplification, and these were correlated with the extent of cytoplasmic accumulation of the constructs. Increased accumulation of GUS/VPg in the cytoplasm was correlated with reduced amplification of the viral genome (Schaad *et al.*, 1996). Localization to the nucleus may down-regulate VPg-Pro functions in the cytoplasm, or the VPg may influence host functions such as transcription by interacting with nuclear factors or structures (Schaad *et al.*, 1996).

Translocation of TEV replicase was very sensitive to changes in the amino acid sequence, and it was proposed that the NLS of the NIb protein is located on the protein surface and mutations that result in alterations of protein folding, inhibit translocation into the nucleus (Li and Carrington, 1993).

Coat Protein

The coat protein encapsidates the viral genome, It has a highly conserved core domain which is necessary for encapsidation and surface-exposed amino and carboxy terminal regions. The amino terminal region varies in length and sequence and is immunodominant (Shukla *et al.*, 1994) over the usually shorter carboxy terminal. The size variation of the coat protein is mostly due to the size variation of the amino terminal region. Important sequence characteristics of the amino terminal region are a DAG sequence important for aphid transmission and the presence of lysine rich sequences (Shukla *et al.*, 1995). The core region has three amino acids, serine, arginine and asparagine, that are highly conserved in filamentous and rod-shaped viruses. Arginine and asparagine residues have been postulated to be part of a salt bridge important for conservation of the capsid protein structure (Dolja *et al.*, 1991).

Coat protein amino acid sequence data has been used to differentiate potyvirus species from potyvirus strains. A bimodal distribution of sequence identity between species and strains was found. Sequence identity between strains of the same species varied from 90 to 99%, whereas identity between different species varied from 38 to 71% (Shukla and Ward, 1988). This finding was inconsistent with a continuum hypothesis of variation in sequences. Further pairwise sequence comparisons of the core regions of the coat proteins resulted in a more complex pattern of distribution consisting of four levels of

identities. Genera show 20 to 30% sequence identity, species within each genus show 55 to 75% sequence identity, subsets of species show sequence identity of 74 to 88% and strains show sequence identity of 90-99% (Shukla *et al.*, 1994).

The primary function of the coat protein is the encapsidation of the viral genome. Recognition of the DAG triplet is involved in transmission by aphid vectors (Atreya *et al.*, 1995) and also seems to be involved in cell to cell and long distance transport of the virus in plants (Dolja *et al.*, 1995). Amino acid changes that resulted in a change in the net charge of the coat protein amino terminal region affected virus movement (Lopez-Moya and Pirone, 1998).

The coat protein of potyviruses is the most important region of the virus genome for the identification and classification of potyviruses. This protein is generally easy to purify, and because of the position of the gene relative to the poly (A) tail it is relatively easy to clone and sequence. The coat proteins of most potyviruses are strongly immunogenic and using the purified coat protein or the expressed coat protein, antisera have been produced to most definite members and some possible members (Shukla *et al.*, 1994). These antisera have been used in a variety of serological techniques for the identification of potyviruses and the study of antigenic relationships between viruses in the family.

Research Objectives

Three potyviruses commonly infect cucurbits in Florida, USA: PRSV-W, ZYMV and watermelon mosaic virus 2 (WMV-2) (Kucharek and Purcifull, 1997; Purcifull *et al.*,

1984). In 1990, a potyvirus with distinct biological, cytological and serological properties was isolated from watermelon in Hendry County, Florida, and was later tentatively named watermelon leaf mottle virus (WLMV) (Purcifull *et al.*, 1998).

Cucurbit crops are grown throughout the world under a variety of ecological conditions and are staple food crops in many parts of the world. Many potyviruses have been described infecting cucurbit plants and eight have been well characterized (Lecoq *et al.*, 1998). Partial or full-length sequences of potyviruses infecting cucurbits available to date are those of several different strains of PRSV-P and -W, WMV-2 and ZYMV. Many of the potyviruses described in the literature are tentative species in the *Potyvirus* genus, and several have not been listed in the 6th Report of the International Committee on Taxonomy of Viruses (ICTV). Some newly described cucurbit potyviruses may be strains of an already named species. Sequence information of the 3' terminal region of the potyviral genome has been used in defining the taxonomic status of the viruses within the genus (Shukla *et al.*, 1994).

The deduced amino acid sequences of potyviral polyproteins show conserved domains, some of which have been experimentally associated with common functions among the different potyviruses studied. However, only a relatively small number of potyviruses have been fully sequenced and most of the studies relating sequence to function have been performed using TEV, TVMV and PPV.

The primary purpose of this study was to perform the molecular and serological characterization of WLMV. The specific objectives were:

1. To obtain clones representing the entire genome of WLMV and to determine the genomic nucleotide sequence and the deduced amino acid sequence of the viral genes.

2. To perform sequence comparisons and phylogenetic analysis of WLMV in order to determine the taxonomic status of WLMV and its relationship to other potyviruses.

3. To identify consensus sequences and domain sequences unique to WLMV and to correlate the sequences to function.

4. To express the coat protein of WLMV in *Escherichia coli*, to produce polyclonal antisera to the expressed coat protein, and to characterize the polyclonal antisera for use in diagnosis and identification of WLMV and for serological relationship studies of WLMV to other potyviruses.

5. To study the host range of WLMV and to identify suitable indicator hosts.

CHAPTER 2
MOLECULAR CHARACTERIZATION OF THE COAT PROTEIN OF WATERMELON LEAF MOTTLE VIRUS (WLMV)

Introduction

Three well characterized potyviruses, i.e., papaya ringspot virus type W (PRSV-W), zucchini yellow mosaic virus (ZYMV) and watermelon mosaic virus 2 (WMV-2), infect cucurbits in Florida, USA (Kucharek and Purcifull, 1997; Purcifull *et al.*, 1984). In 1990, a watermelon sample collected in Florida was discovered to be infected by a *Potyvirus* isolate with distinct biological and serological properties. This isolate was initially referred to as field collection 2932 (FC 2932) and was tentatively named watermelon leaf mottle virus (WLMV) (Purcifull *et al.*, 1998). WLMV is nonpersistently transmissible by the aphid *Myzus persicae* and induces mottle or mosaic in watermelon (*Citrullus lanatus*) and mosaic in squash (*Cucurbita pepo*). It systemically infects several cultivars of cucumber (*Cucumis sativus*) and cantaloupe (*Cucumis melo*) without inducing symptoms, and causes mosaic in 'Prelude II' (Purcifull *et al.*, 1998), a commercial variety of squash with coat protein gene mediated resistance to WMV-2 and ZYMV. WLMV induces the production of conspicuous inclusion bodies in infected plant cells: cylindrical cytoplasmic inclusions, amorphous inclusions in the cytoplasm and nucleus, and elongated or globular inclusions in the nucleus (Petersen *et al.*, 1996; Purcifull *et al.*, 1998). WLMV was serologically distinct from PRSV-W, WMV-2, ZYMV, zucchini yellow fleck virus

(ZYFV) and Moroccan watermelon mosaic virus (MWMV) in reciprocal immunodiffusion tests (Purcifull *et al.*,1998).

In this chapter, we report the nucleotide and the deduced amino acid sequence of the 3' terminal region of WLMV. The relationship of WLMV to 17 other viruses within the genus was determined by sequence comparisons and by phylogenetic analysis of the coat protein and 3' untranslated region.

Materials and Methods

cDNA Library of 3'terminal Region

A cDNA library to WLMV generated by Dr. E. Hiebert using the "Time Saver" cDNA synthesis kit (Pharmacia Biotech Inc., Piscataway, NJ) according to manufacturer's instructions from purified viral RNA, and random and oligo (dT) primers. Blunt-ended cDNA fragments were ligated to EcoR I/Not I adapters and inserted into EcoR I sites of Lambda Zap II (Stratagene, La Jolla, CA). The ligated Lambda DNA was packaged using the "Packagene" Lambda DNA Packaging system (Promega, Madison, WI) (De Sa *et al.*, 1996). Library lysates were used in the subsequent experiments.

Immunoscreening of WLMV Clones

The cDNA library was amplified and clones expressing the coat protein were immunoscreened according to the "Lambda Zap II Instructional Manual" (Stratagene). Positive clones were plaque-purified by three rounds of immunoscreening. Fusion proteins were detected using a rabbit polyclonal antiserum (AS 1163) prepared to the coat protein obtained from dissociated purified virus (Purcifull *et al.*, 1998).

Aliquots of the amplified recombinant library were used to inoculate 200 µl of *Escherichia coli* XL 1 Blue MRF' cells resuspended in 10 mM $MgSO_4$ (OD_{600} = 0.5), and inoculated cells were incubated for 15 min at 37°C. Three ml of melted NZY top agar (0.5% NaCl, 0.2% $MgSO_4.7\ H_2O$, 0.5% yeast extract, 1% casein hydrolysate, and 0.8% agar) were added to the tubes and mixed. This mixture was poured onto NZY plates (0.5% NaCl, 0.2% $MgSO_4.7\ H_2O$, 0.5% yeast extract, 1% casein hydrolysate, and 1.5% agar) and allowed to set. Plates were incubated at 42°C for 4 h until plaques were visible. A nitrocellulose membrane (0.45 µ pore size, Bio-Rad Laboratories, Hercules, CA) was saturated in a 10 mM solution of the gratuitous inducer isopropyl - β-thiogalactopyranoside (IPTG), allowed to air dry, and laid on each plate to induce the expression of the β - galactosidase fusion protein. Plates were then incubated overnight at 37°C. Plates were cooled at 4°C for 30 min, and membranes were removed and washed for 10 min three times in TBST (20 mM Tris-HCl, pH 7.5, 150 mM NaCl, 0.05% Tween-20). Protein binding sites were blocked by incubating membranes for 30 min in Blotto (5% nonfat dry milk in TBST). *E. coli* lysate (Huynh *et al.*, 1984) was added to Blotto at a concentration of 1 mg/ml and membranes were incubated for another 30 min. Membranes were washed for 10 min three times in TBST, followed by incubation overnight at 4°C in the primary antibody, AS 1163 diluted 1:2,000. Membranes were rinsed in TBST for 10 min three times, and were incubated for 2 h at room temperature (RT) in goat anti-rabbit IgG in alkaline phosphatase conjugate (Sigma Chemical Co., St. Louis, MO) diluted 1:30,000 in Blotto. Incubation in the secondary antibody was followed by three washes in TBST, and one in developing buffer (0.1 M Tris-HCl, 0.1 M NaCl, pH 9.6) for 10 min

each. The developing solution was prepared by adding the following reagents to 20 ml of developing buffer: 22 µl of a 75 mg/ml solution of nitroblue tetrazolium (NBT), 20 µl of a 50 mg/ml solution of 5-bromo-4-chloro-3-indolylphosphate (BCIP), and 50 µl 2 M $MgCl_2$. The membranes were incubated in 6 ml of the developing solution containing the color substrates NBT and BCIP (Life Technologies, Gaithersburg, MD), and color development was interrupted after 20 to 30 min, by discarding the developing solution and adding water to the membranes.

Excision of Positive Clones

Positive clones were plaque-purified by immunoscreening three times. Each positive plaque was collected with a sterile Pasteur pipette, and transferred into a microfuge tube containing 100 µl of SM buffer (100 mM NaCl, 50 mM Tris-HCl, pH 7.5, 10 mM $MgCl_2$ and 0.01% gelatin) and three drops of chloroform. Tubes were vortexed and either incubated at RT for 2 h, or overnight at 4°C, to allow for diffusion of phage into the buffer.

Aliquots of the phage solution were added to 10 ml tubes containing 200 µl of XL-1 Blue MRF' cells grown to mid-log phase and resuspended to OD_{600} = 1 in 10 mM $MgSO_4$. The tubes were incubated for 15 min at 37°C for adsorption of phage to bacterial cells. Three ml of LB broth were then added to the tubes and incubated for 3 h at 37°C. Bacterial cells were precipitated by centrifugation at 1,000 *g*, the supernatants were transferred to Oakridge tubes and heated to 70°C for 15 min in a waterbath. The Oakridge tubes were centrifuged at 1,000 *g*, the supernatants were transferred to a fresh tube and

chloroform was added to a final concentration of 5%. The supernatants containing the excised phagemids were either used immediately or stored at 4°C.

Aliquots of 100 µl of each supernatant containing the excised phagemid were added to microfuge tubes containing 200 µl of *E.coli* SOLR cells grown to mid-log phase and resuspended in 10 mM $MgSO_4$ to OD_{600} = 1. The tubes were incubated at 37°C for 15 min and 40 µl of the transformed SOLR cells were plated onto LB/AMP plates (1.0% tryptone, 0.5% NaCl, 0.5% yeast extract, 1.5% agar and 50 µg/ml ampicillin). The plates were incubated overnight at 37°C. For isolation of single colonies, bacterial mass was streaked onto a new plate, and at least two isolated colonies were amplified for clone analysis.

Analysis of WLMV Clones

Individual colonies were grown in LB broth (1% tryptone, 0.5% yeast extract, 0.5% NaCl) with 50 µg/ml ampicillin, at 37°C overnight. Plasmid DNA was purified by alkaline lysis miniprep as described by QIAGEN Inc. (Valencia, CA). Cells from overnight cultures of 3 to 4 ml were pelleted by centrifugation at 2,000 *g* for 10 min. Pellets were resuspended in 300 µl of P1 buffer (50 mM Tris-HCl, pH 8.0, 10 mM EDTA, RNAse A 100 µg/ml), and were incubated for 5 min on ice. Buffer P2 (200 mM NaOH, 1% SDS) was freshly prepared and 300 µl were added to each tube, the tubes were gently mixed by inversion, and 300 µl of buffer P3 (3 M potassium acetate, pH 5.5) were added. The tubes were incubated on ice for 10 min, followed by centrifugation at 16,000 *g* for 10 min at 4°C. The supernatants were collected and mixed with one volume of isopropanol, and plasmid DNA was precipitated for 30 min at RT. The DNA was pelleted by centrifugation

at 16,000 *g* at 4°C for 20 min, and the pellet was washed with 70% ethanol and vacuum dried. The DNA was resuspended in 30 µl of molecular biology grade water (MBG), and screened for correct insert size by digestion with EcoR I (Life Technologies, Gaithersburg, MD).

Sequence of 3'terminus Clones

Two clones that were positive in immunoscreening and of adequate size, judging by restriction enzyme digestion with EcoR I, were initially sequenced by the dideoxy chain termination method (Sanger *et al.*, 1977). The "Sequenase Version 2.0 DNA T7 Polymerase" sequencing kit (United States Biochemical, Cleveland, OH), [^{35}S] dATP, and primers T7 and T3 complementary to the pBlueScript vector, were used. Sequencing runs were performed in 6% polyacrylamide gels containing 7 M urea. Autoradiographs of air-dried sequencing gels were made using XAR X-ray film (Eastman Kodak, Rochester, NY).

All subsequent sequencing was done by automated sequencing at the DNA Sequencing Core Facility of the Interdisciplinary Center for Biotechnology Research (ICBR) at the University of Florida. Two overlapping clones (pCPA and pCPB) and two EcoR I restriction site-directed subclones of clone pCPB were sequenced using M13 forward and reverse primers. Two internal sequencing primers, 5' GGGTCTGGACTATGAGGA 3' and 5' ACAGAACAAAACGAATTCCTT 3', were designed and used to confirm the reading frame and the cleavage site between the replicase and the coat protein.

Sequence Analysis

Analyses of nucleotide and predicted amino acid sequences were done using the Wisconsin Package Version 8.1 (Genetics Computer Group [GCG], Madison, WI) available at the ICBR Biological Computing Facility, University of Florida. Sequences of 17 potyviruses were chosen from the output of the Basic Local Alignment Search Tool (BLAST) (Altschul *et al.*, 1990) and obtained from GenBank. Sequences were chosen according to the following criteria: the ten highest BLAST scores, viruses infecting cucurbits, viruses representing subgroups or of economical importance, and the sequence with the lowest BLAST score. Pairwise percentage identities were calculated using GAP in Version 10 of the Wisconsin Package.

CLUSTAL W (1.7) was used to align the sequences and to construct neighbor-joining trees calculated by the method of Saitou and Nei as implemented in CLUSTAL W (1.7) (Thompson *et al.*, 1994). Reliability of the trees was estimated using bootstrap with 1,000 replicates. Trees generated by CLUSTAL W (1.7) were viewed using TreeView 1.2 (Page, 1996). Multiple sequence alignments were also made using PILEUP, and phylogenetic trees were generated by maximum parsimony using PAUP 3.1.1 (Swoford, 1993). The optimal trees were obtained by a heuristic search with the tree bissection-reconnection option. The reliability of internal branches of the most parsimonious trees was assessed using the bootstrap method with 100 replications as implemented in PAUP.

Results

Two overlapping clones of the 3'terminal region of WLMV ending at the poly (A) tail were selected after initial sequencing by the dideoxy chain termination method: clone pCPA with 1,085 nucleotides (nt) and clone pCPB with 2,300 nt, totaling 2,316 nt (Fig. 2-1). The two EcoR I directed subclones of clone pCPB were pCPBl with 1,537 nt and pCPBs with 763 nt. Sequencing of clones pCPA and pCPB and of the two EcoR I directed subclones of pCPB revealed an open reading frame (ORF) of 2,050 nt ending at an ochre (UAA) termination codon (Fig. 2-1; Fig. 2-2). Clones pCPA and pCPB included part of the region encoding the replicase and the entire coat protein coding region, followed by 243 nt of the 3'UTR and 23 adenosine residues of the poly (A) tail (GenBank Accession no. AF028004). The predicted amino acid sequence encompassed 398 amino acids (aa) of the replicase and 284 aa of the coat protein with a calculated size of 32.6 kDa (Fig. 2-2).

The consensus motif (T/S)GXXXTXXXNS(T/S) (921 to 38 aa) GDD, which is conserved in RNA-dependent RNA polymerases (Domier *et al.*, 1987), was identified in the replicase with the sequence SGQOSTVVDNSL (29 aa) GDD. A putative cleavage site with the sequence VHFQ/S was found between the replicase and the coat protein genes. This is in agreement with the group specific motif VXXQ/S, A, V, G suggested for the VPg-protease (VPg-Pro) cleavage sites (Shukla *et al.*, 1991). A DAG amino acid triplet (Atreya *et al.*, 1995) was found 25 aa downstream of the presumed protease cleavage site within the amino terminal domain of the coat protein. Trypsin cleavage sites were identified at the boundary between the conserved core region and the amino terminal

region (Shukla *et al.*, 1988). Within the core region, there were three amino acid residues highly conserved in filamentous and rod-shaped viruses: a serine (S) residue at position 152, and arginine (R) at position 184, and an aspartate (D) at position 218 (Fig. 2-2). The two latter residues have been suggested to be important for assembly or stability maintenance of the virus particles (Dolja *et al.*, 1994).

The 3'UTR of WLMV was determined to be 243 nt long, with a nucleotide composition of A:50, C:53, G:58 and U:82. The sequence of the 3'UTR was made up of several short repeats (three CACACC, six GAUA and four other repeats of four bases) and was predicted to fold into a stable secondary conformation using mfold Version 3.0 (Zuker, 1989).

Coat protein sequence identities ranging from 54.8 to 65.2% were found between WLMV and 17 other members of the *Potyvirus* genus at the amino acid level and from 56 to 64% at the nucleotide level. Percentage of sequence identity among 3'UTR ranged from 37% to 48.3% (Table 2-1).

Phylogenetic trees of coat protein amino acid sequences and nucleotide sequences of the 3'UTR generated by CLUSTAL W (1.7) are shown in Fig. 2-3 and in Fig. 2-4, respectively.

The bootstrap 50% majority rule consensus tree generated by PAUP of coat protein amino acid sequences had three highly supported branches. Bootstrap support of 99% was found for a branch with PeSMV and PVY-T, and of 84% for a BCMV subgroup branch containing BCMV, WMV-2 and ZYMV-CA. A branch with MDMV and OYDV had 81% bootstrap support.

The bootstrap 50% majority rule consensus tree of the sequence of the 3'UTR had a branch with 100% bootstrap support with BCMV and WMV-2, but not ZYMV-CA. Other branches were a PeSMV and PVY-T branch with 96% support and an OYDV and TuMV branch with 54% support.

Phylogenetic analysis using PAUP indicated that it was not possible to infer phylogenetic relationships between WLMV and the other viruses using this set of data. WLMV did not cluster in a clade with high bootstrap support in 50% majority rule consensus trees.

Discussion

The nucleotide sequence of the 3'terminal region of WLMV and the deduced amino acid sequence of part of the replicase and the coat protein have been determined, and consensus motifs typical of potyvirus sequences have been identified.

Percentage of sequence identity between coat protein amino acid sequences of WLMV and of 17 other potyviruses, are within the range of 38 to 71% identity found for different potyvirus species, according to the theory of bimodal distribution (Shukla and Ward, 1988). Nucleotide sequence comparisons of the 3'UTR are within the range of sequence identities of 30 to 53% found for the 3'UTR of different species of potyviruses (Frenkel *et al.*, 1989). The highest percentage of amino acid sequence identity was 65.2% between WLMV and PeSMV. Sequence identities between the coat proteins of WLMV and the coat proteins of the potyviruses most commonly infecting cucurbits, PRSV-W, WMV-2 and ZYMV were 59.6%, 56.9% and 56.7%, respectively. The percentage of

sequence identities between WLMV and OrMV and OYDV were 62.8 and 60.5%, respectively. The alignment of coat protein sequences in Fig. 2-5 show that OYDV and OrMV have short amino termini of 21 and 23 aa, respectively, therefore the percentage of sequence identity is based primarily on the highly conserved core.

Sequence identity between the coat protein amino acid sequence of PRSV-W and WLMV, is 59.6% and between the nucleotide sequences of the 3'UTR is 38.9%. Like PRSV-W, the amino terminal region of the coat protein of WLMV is lysine enriched, and the repeat KEKEK is present in both viruses, once in WLMV and three times in PRSV-W (Fig. 2-5). Lysine rich amino termini are found in some potyviruses and may be of evolutionary significance as suggested previously (Hammond, 1992). Partial gene duplications resulting in repeats at the amino terminal region of the coat protein have been suggested to be an important evolutionary mechanism for generation of *Potyvirus* species with new host specificities (Xiao *et al.*, 1993). Additionally, the KEKEK repeat could be biologically important as a determinant of symptom severity or host range.

The overall topology of the trees obtained using maximum parsimony and the neigbor-joining distance matrix methods was similar as far as the branches with high bootstrap support are considered. The highly supported branches are consistent with a previous report (Berger *et al.*, 1997), and are indications of the robustness of the trees generated by both methods.

Inconsistencies in the positions of taxa using different methods of phylogenetic analysis suggest that phylogenetic trees based on small numbers of divergent sequences may lead to erroneous assumptions about evolutionary relationships. Branches with low bootstrap support are, as indicated previously (Berger *et al.*, 1997), more likely to be in

disagreement, and high frequency of such branches yields uninformative trees from which phylogenetic inferences cannot be made.

WLMV did not cluster with other viruses in branches with high bootstrap support in the phylogenetic trees, and no evidence was found for its inclusion in a subgroup with any of the other viruses. Comparative sequence analysis show that WLMV is a unique species within the *Potyvirus* genus.

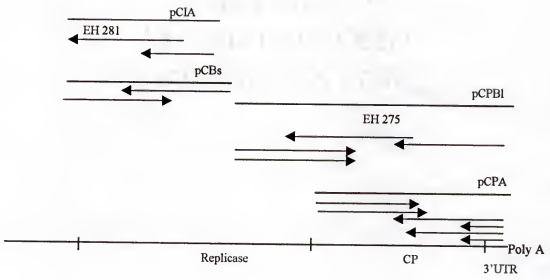

Fig. 2-1. Sequencing strategy used for the cDNA clones representing the 3'terminal region of watermelon leaf mottle virus (WLMV). Arrows indicate the direction of sequencing and distance read from the beginning of the clone or from the location of a primer (EH 275 and EH 281).

Fig. 2-2. Nucleotide sequence of the 3'terminal region of watermelon leaf mottle virus (WLMV). The predicted amino acid (aa) sequence of the open reading frame for the putative polyprotein is shown. Underlined amino acids correspond to those of the replicase consensus motif SGQPSTVVDNSL (29 aa) GDD (aa 189-230), the genome-linked protein protease VHFQ/S cleavage site between the replicase and the coat protein (aa 395-399), and the DAG aphid transmissibility motif (aa 424-426). A trypsin cleavage site at aa residues 446-453 is also underlined. Conserved amino acids in the core region of potyviruses are in bold and italics: serine (aa 541), arginine (aa 573) and aspartate (aa 617).

```
1     CTTGAATATGAACGCAGCAGTGGGATCATTGTTTGTTGGCAAGAAACGAGCGTACTTCGA 60
        L  N  M  N  A  A  V  G  S  L  F  V  G  K  K  R  A  Y  F  E

61    AACATGCACAGCGCAAGATAGAGAAGATCTAGTTCAAGCCAGCTGTAAGCGTCTATATTA 120
        T  C  T  A  Q  D  R  E  D  L  V  Q  A  S  C  K  R  L  Y  Y

121   TGGTGAAAAAGGAGTCTGGAATGGATCCTTGAAAGCTGAATTGAGACCTATTGAGAAAGT 180
        G  E  K  G  V  W  N  G  S  L  K  A  E  L  R  P  I  E  K  V

181   TGAATTAAAGAAAACTCGTACTTTCACAGCGGCACCCATCGATACTTTACTTGGTGGAAA 240
        E  L  K  K  T  R  T  F  T  A  A  P  I  D  T  L  L  G  G  K

241   AGTGTGTGTTGATGATTTCAACCAGAAATTTTATAACCACCACATTAAAGGTCCGTGGAC 300
        V  C  V  D  D  F  N  Q  K  F  Y  N  H  H  I  K  G  P  W  T

301   CGTCGGGATGACGAAGTTCTACAAAGGGTGGGACGAACTATTAAACTGTCTGCCGGACAA 360
        V  G  M  T  K  F  Y  K  G  W  D  E  L  L  N  C  L  P  D  N

361   CTGGATCTATTGCGATGCCGATGGGAGCCAATTCGACAGTTCGATCACACCATACATCAT 420
        W  I  Y  C  D  A  D  G  S  Q  F  D  S  S  I  T  P  Y  I  M

421   GAATGCAGTTTTGAAGGTACGTGAGCATTTTATGGAAGACTGGGACCTTGGGAAGGAGAT 480
        N  A  V  L  K  V  R  E  H  F  M  E  D  W  D  L  G  K  E  M

481   GTTACGAAATCTTTACACTGAGATAGTGTATACGCCTATATCTACACCAGATGGCACGAT 540
        L  R  N  L  Y  T  E  I  V  Y  T  P  I  S  T  P  D  G  T  I

541   AATCAAGAAACACAAAGGGAACAATAGTGGCCAACCATCTACAGTCGTAGATAATTCACT 600
        I  K  K  H  K  G  N  N  S  G  Q  P  S  T  V  V  D  N  S  L

601   AATTGTGTGCATCGCTGTGAAATATGCGCTCTTAAAAGAAGGATTGAAACCAGAAGAGTT 660
        I  V  C  I  A  V  K  Y  A  L  L  K  E  G  L  K  P  E  E  F

661   TGATAACGTGATCCGTTTCTTTGTTAATGGAGACGATCTTCTCATAGCCGTGAAACCATC 720
        D  N  V  I  R  F  F  V  N  G  D  D  L  L  I  A  V  K  P  S

721   CCATGAGCACATCTTGAACAACTTCTCAGAATACTTCGCTGAGCTTGGATTGAACTACGA 780
        H  E  H  I  L  N  N  F  S  E  Y  F  A  E  L  G  L  N  Y  E

781   ATTCGACAACCGACACACTGACAAATCCAGACTTTGGTTTATGTCCCACTGTGGCATTCG 840
        F  D  N  R  H  T  D  K  S  R  L  W  F  M  S  H  C  G  I  R

841   TGATGAAGATATCTGGATCCCGAAACTTGAACCCGAACGCATAGTGTCTATCCTTGAATG 900
        D  E  D  I  W  I  P  K  L  E  P  E  R  I  V  S  I  L  E  W

901   GGATAGAAGTAAATTGCCTGAGAATCGTCTGGAAGCAGTGTGTGCCGCCATGGTTGAGGC 960
        D  R  S  K  L  P  E  N  R  L  E  A  V  C  A  A  M  V  E  A

961   GTGGGGCTATAATGATTTACTTCACCAAATACGGTTGTTTTACCAATGGCTCCTCGAACA 1020
        W  G  Y  N  D  L  L  H  Q  I  R  L  F  Y  Q  W  L  L  E  Q

1021  AGAGCCATTCAACGATCTGGCCGCAAATGGGTATGCACCGTACATTAGTGAAGTAGCATT 1080
        E  P  F  N  D  L  A  A  N  G  Y  A  P  Y  I  S  E  V  A  L

1081  GCGCGCGCTGTATACAGGAAGCCAACCAACCAATACATCAATGCTACCATATTTACATCA 1140
        R  A  L  Y  T  G  S  Q  P  T  N  T  S  M  L  P  Y  L  H  Q

1141  ACTTGTCGATGATTTTGAACATGATGCAGACGAAGATTATGAAGTTCATTTCCAGAGTGG 1200
        L  V  D  D  F  E  H  D  A  D  E  D  Y  E  V  H  F  Q  /S  G
```

```
1201 TGAAAAAGAAAAGGAGAAAATTGATACGGAGGAGGCTAAGCGTACAATAAAGAAATTGAA 1260
       E  K  E  K  E  K  I  D  T  E  E  A  K  R  T  I  K  K  L  K

1261 AGAGGTTGTCGATGCAGGCAATGAGCAATCTATTACTACATCCTCGAAAGTTAAGCAGGA 1320
       E  V  V  D  A  G  N  E  Q  S  I  T  T  S  S  K  V  K  Q  D

1321 CCAAATAAGTACCGAAAAGGACAAAGATGTTTCAACCTCGACGCGTGGCACATTTGCAAT 1380
       Q  I  S  T  E  K  D  K  D  V  S  T  S  T  R  G  T  F  A  I

1381 TCCAAAACTGAGATTGATGAATCAAGTCATGAAACTACCAAAGATTGGAAATAGAAATTT 1440
       P  K  L  R  L  M  N  Q  V  M  K  L  P  K  I  G  N  R  N  L

1441 ATTAAATCTCGAACACTTACTCACGTATAAGCCTGAACAAGTTGACATATCCAATACGCG 1500
       L  N  L  E  H  L  L  T  Y  K  P  E  Q  V  D  I  S  N  T  R

1501 CGCAACAAAGCAACAGTTTGAAACTTGGTGTGAACGTGTTCGTGATGCGTATGGATTGAA 1560
       A  T  K  Q  Q  F  E  T  W  C  E  R  V  R  D  A  Y  G  L  K

1561 AGATGATGAAATGAGTATAGTGATGAATGGATTTATGGTCTGGTGCATTGAAAATGGCAC 1620
       D  D  E  M  S  I  V  M  N  G  F  M  V  W  C  I  E  N  G  T

1621 ATCACCCAATGTCAATGGGGTCTGGACTATGATGGATGGCACAGAACAAAACGAATATCC 1680
       S  P  N  V  N  G  V  W  T  M  M  D  G  T  E  Q  N  E  Y  P

1681 TTTGAAACCCATGGTTGAAAATGCGAAACCTACCCTCCGCCAAATCATGGCGCATTTCTC 1740
       L  K  P  M  V  E  N  A  K  P  T  L  R  Q  I  M  A  H  F  S

1741 TGACGCTGCGGAAGCTTACATTGAAATGCGGAACACAAAGGAACCATATATGCCAAGATA 1800
       D  A  A  E  A  Y  I  E  M  R  N  T  K  E  P  Y  M  P  R  Y

1801 TGGACAGCAGCGTAACCTAACCGACATGAGCCTTGCACGATGCGCTTTCGATTTTTATGA 1860
       G  Q  Q  R  N  L  T  D  M  S  L  A  R  C  A  F  D  F  Y  E

1861 AATGACATCTCGAACACCAGCACGTGTCCGAGAAGCTCATATTCAAATGAAAGCAGCTGC 1920
       M  T  S  R  T  P  A  R  V  R  E  A  H  I  Q  M  K  A  A  A

1921 TCTGCGTAATACAAGAACTAGAATGTTTGGACTTGATGGCAATGTTGGCACTCAAGAGGA 1980
       L  R  N  T  R  T  R  M  F  G  L  D  G  N  V  G  T  Q  E  E

1981 GGATACTGAGAGACACACCACTGAGGACGTTAATAGGAACATGCATAATCTTCTTGGTCT 2040
       D  T  E  R  H  T  T  E  D  V  N  R  N  M  H  N  L  L  G  L

2041 CAAAATGTAATACTATATAATGCATGTGTGGTGAACTATCCTATATATAGCAGTATCTGT 2100
       K  M

2101 ATGTTTATTCATCTCCTCTCCCTTGCAGACCCTGATATCCTGTAGTGTGGGTGGCCCACC 2160

2161 AGCGCTTAGGCAGCCAGGACAGGTTCTATTCTGTAAGTGCTCTTCTGCTATGTTGCCTAG 2220

2221 GTTAGGAGTAAAAGTCTATCGCATGCGTGTGGTAGGCTTCCCGCGATCCTATGTTTCATA 2280

2281 GTAGCTTTGAGTGAAAAAAAAAAAAAAAAAAAAAA 2316
```

Table 2-1. Percentage of sequence identities between the nucleotide (nt) and amino acid (aa) sequences of the coat protein, and the 3' untranslated region (3'UTR) of watermelon leaf mottle virus (WLMV) (GenBank Accession no. AF028004), and 17 viruses of the *Potyvirus* genus.

VIRUS [a]	COAT PROTEIN nt %	COAT PROTEIN aa %	3' UTR nt %	Accession numbers
OrMV	63.5	62.8	38.8	dbj D00615
PeSMV	63.3	65.2	42.3	emb X66027
PVY-T	63.3	63.3	37.0	dbj D12570
BYMV	63.2	57.1	43.4	dbj D00490
TuMV	62.6	58.0	36.8	gb U25663
OYDV	62.4	60.5	37.4	dbj D73378
LYSV	62.3	58.1	39.1	dbj D11118
TBV	62.2	62.9	48.3	emb X63630
ZYMV-CA	61.3	56.7	45.0	gb L31350
PPV	61.2	58.1	40.6	emb X56258
LMV	61.2	56.2	42.8	gb U24670
MDMV-A	60.0	57.2	37.2	gb U07216
TEVn	59.7	58.8	44.2	gb L38714
BCMV	59.4	54.8	44.2	emb Z15057
PRSV-W	59.1	59.6	38.9	dbj D00594
WMV-2	59.1	56.9	39.8	dbj D00535
SwPLV	55.6	55.1	42.0	emb Z26136

[a] Ornithogalum mosaic virus (OrMV), pepper severe mosaic virus (PeSMV), T strain of potato virus Y (PVY-T), bean yellow mosaic virus (BYMV), turnip mosaic virus (TuMV), onion yellow dwarf virus (OYDV), leek yellow stripe virus (LYSV), tulip breaking virus (TBV), California strain of zucchini yellow mosaic virus (ZYMV-CA), El Amar strain of plum pox virus (PPV), lettuce mosaic virus (LMV), strain A of maize dwarf mosaic virus (MDMV-A), necrotic strain of tobacco etch virus (TEVn), bean common mosaic virus (BCMV), papaya ringspot virus type W (PRSV-W), watermelon mosaic virus 2 (WMV-2), and sweet potato latent virus (SwPLV). Sequences were obtained from GenBank.

Fig. 2-3. Relationship between selected potyviruses, based on coat protein amino acid sequences using the neighbor-joining method as implemented in CLUSTAL W (1.7). Numbers represent bootstrap values with 1000 replications.
Watermelon leaf mottle virus (WLMV), Ornithogalum mosaic virus (OrMV), pepper severe mosaic virus (PeSMV), T strain of potato virus Y, (PVY-T), bean yellow mosaic virus (BYMV), turnip mosaic virus (TuMV), onion yellow dwarf virus (OYDV), leek yellow stripe virus strain (LYSV), tulip breaking virus (TBV), California strain of zucchini yellow mosaic virus, (ZYMV-CA), El Amar strain of plum pox virus (PPV), lettuce mosaic virus (LMV), strain A of maize dwarf mosaic virus (MDMV-A), tobacco etch virus necrotic strain (TEVn), bean common mosaic virus (BCMV), papaya ringspot virus type W (PRSV-W), watermelon mosaic virus 2 (WMV-2), sweet potato latent virus (SwPLV). Sequences were obtained from GenBank. The value on the scale bar means 0.1 nucleotide substitutions per site.

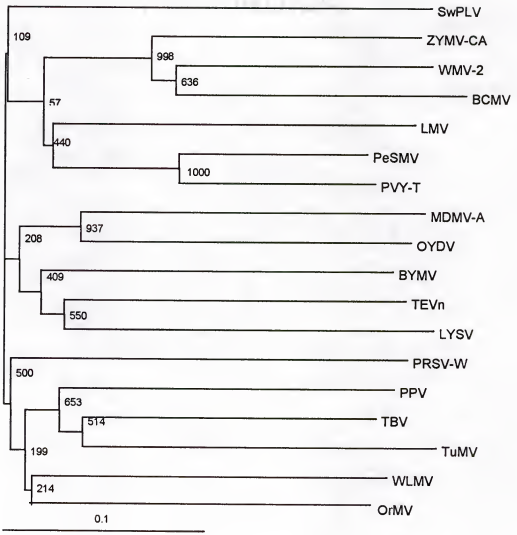
SwPLV
109
ZYMV-CA
998
WMV-2
636
BCMV
57
LMV
440
PeSMV
1000
PVY-T
MDMV-A
937
208
OYDV
BYMV
409
TEVn
550
LYSV
PRSV-W
500
PPV
653
514
TBV
199
TuMV
WLMV
214
OrMV
0.1

Fig. 2-4. Relationship between selected potyviruses based on nucleotide sequences of the 3' unranslated region (3'UTR) using the neighbor-joining method as implemented in CLUSTAL W (1.7). Numbers represent bootstrap values out of 1000 replications.
Watermelon leaf mottle virus (WLMV), Ornithogalum mosaic virus (OrMV), pepper severe mosaic virus (PeSMV), T strain of potato virus Y (PVY-T), bean yellow mosaic virus (BYMV), turnip mosaic virus (TuMV), onion yellow dwarf virus (OYDV), leek yellow stripe virus strain (LYSV), tulip breaking virus (TBV), California strain of zucchini yellow mosaic virus (ZYMV-CA), El Amar strain of plum pox virus (PPV), lettuce mosaic virus (LMV), strain A of maize dwarf mosaic virus (MDMV-A), necrotic strain of tobacco etch virus (TEVn), bean common mosaic virus (BCMV), papaya ringspot virus type W (PRSV-W), watermelon mosaic virus 2 (WMV-2), and sweet potato latent virus (SwPLV). Sequences were obtained from GenBank. The value on the scale bar means 0.1 nucleotide substitutions per site.

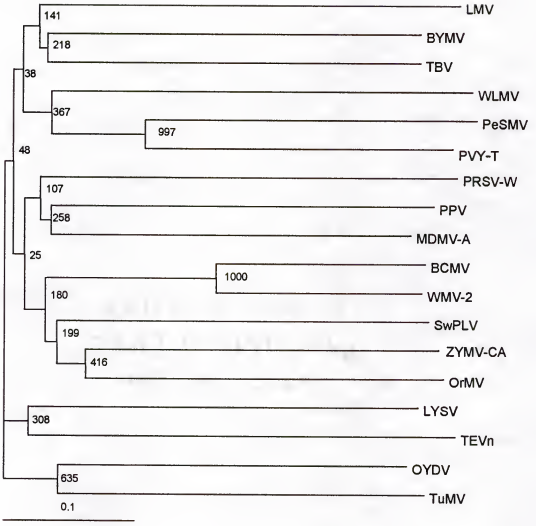
LMV
141
BYMV
218
TBV
38
WLMV
367
PeSMV
997
PVY-T
48
PRSV-W
107
PPV
258
MDMV-A
25
BCMV
1000
180
WMV-2
SwPLV
199
ZYMV-CA
416
OrMV
LYSV
308
TEVn
OYDV
635
TuMV
0.1

Fig. 2-5. Multiple sequence alignment of coat protein amino acid sequences of watermelon leaf mottle virus (WLMV), Ornithogalum mosaic virus (OrMV), papaya ringspot virus type W (PRSV-W), onion yellow dwarf virus (OYDV), watermelon mosaic virus 2 (WMV-2), zucchini yellow mosaic virus (ZYMV). Alignment done using CLUSTAL W (1.7). Asterisks indicate conserved residues and dots indicate conservative substitutions.

```
WLMV        ---------S---------GEKEKEKIDTEEAKRTIKKLKEVVDAGN------EQSITTS
OrMV        ------------------------------------ADSMDAG-------GSSRPPA
PRSV-W      ---------SKNEAVDTGLNEKFKEKEKQKEKEKEKQKEKEKDDASD------GNDVSTS
OYDV        ------------------------------------ASEIEDAAN---------VNTD
WMV-2       -------------------SGKE-AVENLDTGKDSKKDTSGKGDKPQNSQ--TGQGSKEQ
ZYMV-CA     -------------------SGTQPTVADAGATKKDKEDDKGKNKDV------TGSGSGEK
                                                              .

WLMV        SKVKQDQI---STEKDKDVSTSTRGTFAIPKLRLMNQVMKLPKIGNRNLLNLEHLLTYKP
OrMV        PLVRQ---------QDQDVNV---GTFSVARVKALSDKMMLPKVRGKTVLNLQHLVQYNP
PRSV-W      TK---------TGERDRDVNVGTSGTFTVPRIKSFTDKMILPRIKGKSVLNLNHLLQYNP
OYDV        KQVGK------SKDKDKDVDVGTSGEFSVPKVKMLSDKMRLPRVGKKVILNGKHLLTYKP
WMV-2       TKIG---------TVSKDVNVGSKGK-EVPRLQKITKKMNLPTVGGKIILSLDHLLEYKP
ZYMV-CA     TVAAV--------TKDKDVNAGSHGK-IVPRLSKITKKMSLPRVKGNVILDIDHLLEYKP
                          .:**..   *   :.::  :.. * ** :  . :*. .**: *:*

WLMV        EQVDISNTRATKQQFETWCERVRDAYGLKDDEMSIVMNGFMVWCIENGTSPNVNGVWTMM
OrMV        EQTEISNTRATRTQFNNWYDRVRDSYGVTDDQMAVILNGLMVWCIENGTSPNLNGNWTMM
PRSV-W      QQIDISNTRATQSQFEKWYEGVRNDYGLNDNEMQVMLNGLMVWCIENGTSPDISGVWVMM
OYDV        DQVDLYNTRATHAQFKTWYDAVKLEYELTDEQMKIVMNGLMVWCIENGTSQNLTGVWTMM
WMV-2       NQVDLFNTRATKTEFESWYSAVKIEYDLNDEQMGVIMNGFMVWCIDNGTSPDVNGVWVMM
ZYMV-CA     DQIELYNTRASHQQFASWFNQVKTEYDLNEQQMGVVMNGFMVWCIENGTSPDINGVWVMM
            :* :: ****:: :* .* . *:  * :.:::* :::**:*****:**** ::.* *.**

WLMV        DGTEQNEYPLKPMVENAKPTLRQIMAHFSDAAEAYIEMRNTKEPYMPRYGGQRNLTDMSL
OrMV        DGDEQIEYPLQPVLENAQPTFRQIMAHFSNAAEAYIEKRNSEQRYMPRYGSQRNLNDYSL
PRSV-W      DGETQVDYPIKPLIEHATPSFRQIMAHFSNAAEAYIAKRNATERYMPRYGIKRNLTDISL
OYDV        DGHNQMEYPLSPIIDNAKPTFRQIMAHFSYAAEAYIEYRNATEKYMPRYGLQRNLREYSL
WMV-2       DGEEQVEYPLKPIVENAKPTLRQIMHHFSDAAEAYIEMRNSESPYMPRYGLLRNLRDREL
ZYMV-CA     DGNEQVEYPLKPIVENAKPTLRQIMHHFSDAAEAYIEMRNAEAPYMPRYGLLRNLRDRSL
            **  * :**:.*::::* *::**** *** ******  **:   ******  *** : .*

WLMV        ARCAFDFYEMTSRTPARVREAHIQMKAAALRNTRTRMFGLDGNVGTQEEDTERHTTEDVN
OrMV        ARYAFDFYEMTSRTANRAREAHIQMKAAALRNTKTKLFGLDGKVGTEEEDTERHVASDVN
PRSV-W      ARYAFDFYEVNSKTPDRAREAHMQMKAAALRNTSRRMFGMDGSVSNKEENTERHTVEDVN
OYDV        ARYAFDFYEMNSKTPIRAKEAHMQMKAAAVRGVANRMFGLDGNISTDDENTERHTAADVN
WMV-2       ARYAFDFYEVTSKTPNRAREAIAQMKAAALAGINSRLFGLDGNISTNSENTERHTARDVN
ZYMV-CA     ARYAFDFYEVNSKTPERAREAVAQMKAAALSNVSSRLFGLDGNVATTSEDTERHTARDVN
            ** ******:.*:*. *.:**  ******: .   ::**:**.:.. .*:****.. ***

WLMV        RNMHNLLGLKM
OrMV        RNMHSLLGVNM
PRSV-W      RDMHSLLGMRN
OYDV        KDHHTLLGLRM
WMV-2       QNMHTLLGMGPPQ
ZYMV-CA     RNMHTLLGVNTMQ
            :: *.***:
```

CHAPTER 3
CLONING AND COMPLETE NUCLEOTIDE SEQUENCING OF WATERMELON LEAF MOTTLE VIRUS

Introduction

The potyvirus genome is a single-stranded positive sense RNA with a single open reading frame (ORF) translated into a polyprotein that is co- and posttranslationally cleaved into at least nine proteins. The deduced amino acid sequences of potyviral proteins show conserved domains, some of which have been associated with common functions among the different viruses in the genus. However, only a relatively small number of potyviruses have been fully sequenced and most studies associating sequence to function performed to date, have been done with viruses that primarily infect tobacco, i.e. tobacco etch virus (TEV) and tobacco vein mottling virus (TVMV). The determination of the sequence of a virus allows for the identification of consensus regions that may be associated with function. It also enables the construction of infectious clones that can be used to investigate the importance of those identified regions. Computer based predictions of cellular localization and function may be useful in designing experiments to study viral gene functions in virus-host interactions. In general, three dimensional protein structure cannot be predicted from sequence, although it may be predicted by homology modeling (Rost and Valencia, 1996). However, theoretical biology can be used to assist in

predicting unidimensional aspects of three dimensional structures, such as secondary structures, sequence motifs, transmembrane helices, binding sites, solvent accessibility, and some aspects of protein function (Rost and Valencia, 1996).

Nucleic acid comparisons of the full length sequences of viruses in the *Potyvirus* genus show a range of identities from 30 to 60% for distinct species, and a range of identities greater than 95% for strains of a virus species (Shukla *et al.*, 1994). High levels of amino acid sequence identities are also found when strains of the same virus are compared, and much lower levels of identity are found for different species. The greatest variability in amino acid sequence is found within the amino terminal region of the coat protein, followed by the P1 and the P3 proteins. The most conserved protein is the replicase with identity of 61 to 64%, whereas the least conserved proteins are the P1 protein with an identity range of 10 to 19%, and the P3 protein with a range of identities from 30 to 33% (Shukla *et al.*, 1994).

Three potyviruses are found commonly infecting cucurbits in Florida, USA: papaya ringspot virus type W (PRSV-W), zucchini yellow mosaic virus (ZYMV) and watermelon mosaic virus 2 (WMV-2) (Kucharek and Purcifull, 1997, Purcifull *et al.*, 1984). Watermelon leaf mottle virus (WLMV) was collected in Hendry County, Florida, in 1990, and has biological, cytological and serological properties distinct from PRSV-W, ZYMV and WMV-2.

The objectives of the research reported in this chapter were to:

1 - Obtain clones representing the entire genome of WLMV and determine the sequence of the virus.

2 - Compare the full length sequence of WLMV with that of other potyviruses.

3 - Identify consensus regions in the nucleotide and amino acid sequence that may be associated to protein functions.

4 - Perform phylogenetic analysis of the entire amino acid sequence of WLMV.

Materials and Methods

Virus Isolate

Cotyledons of squash seedlings (*Cucurbita pepo*) 'Early Prolific Straightneck' (EPSN) were inoculated with an isolate of WLMV that had been preserved by desiccation over silica gel and stored at 4°C. Mechanical inoculation was performed by grinding EPSN leaf tissues with a mortar and pestle in potassium phosphate buffer (20 mM, pH 7.5) and 600 mesh Carborundum. The inoculated plants were maintained in a growth room at 24°C.

Purification of watermelon leaf mottle virus

WLMV was purified using a modification of the protocol described by Lecoq and Pitrat (1985). All the reagents were used at 4°C. EPSN squash leaves were harvested at 25 days post-inoculation and homogenized for 1 min in a precooled Waring blender with 2.5 volumes of purification buffer (0.3 M K_2HPO_4, pH 8.3, 0.2% sodium diethyldithiocarbamate [DIECA] and 0.1% β - mercaptoethanol [β-ME]). The resulting slurry was emulsified for 1 min with 2/3 volumes of trichlorofluorethane (Freon) and filtered through cheesecloth. The filtrate was centrifuged at 5,000 *g* for 10 min in an SS 34 rotor in a Sorvall centrifuge (Du Pont Co., Wilmington, DE), the supernatant was

collected, and the centrifugation step was repeated. The aqueous phase was collected and Triton X-100 was added to a final concentration of 1% and stirred for 30 min at 4°C. The mixture was centrifuged for 90 min at 4°C at 100,000 *g* in a Beckman Ti 70 rotor (Beckman Instruments, Inc., Palo Alto, CA). The pellet was resuspended in 20 mM Hepes, pH 8.2 for 1 h at 4°C with stirring. The suspension was partially clarified by centrifugation at 2,000 *g* for 10 min. The supernatant was layered onto a Cs_2SO_4 gradient (10 g Cs_2SO_4 in 27 ml 20 mM Hepes, pH 8.2) and centrifuged at 140,000 *g* for 16 h at 4°C, in a Beckman SW 41 rotor. The birefringent band was collected using a syringe, mixed with one volume of 20 mM Hepes, pH 8.2, and centrifuged at 10,000 *g* for 10 min in an SS 34 rotor. The virus was precipitated from the supernatant by addition of a 20% polyethylene glycol (PEG, MW 8,000) solution to a final concentration of 4% PEG (w/v). After incubation on ice for 20 min, the sample was centrifuged at 10,000 *g* for 10 min. The virus pellet was resuspended in 20 mM Hepes, pH 8.2, and virus concentration was estimated by spectrophotometry using an approximate extinction coefficient of $A_{260} = 2.5$ (1 mg/ml, 1 cm light path).

The virus preparation was checked by electron microscopy and by electrophoresis in 10% sodium dodecyl sulphate polyacrylamide gels (10% SDS-PAGE) stained in Coomassie blue. To check the purified virus preparation by electron microscopy, formvar/carbon coated copper grids were floated on a drop of the virus suspension. The grid was washed with distilled water, blotted with Whatman filter paper, stained by floating on one drop of uranyl acetate 2%, and then dried. The purified virus preparation

was divided into aliquots and stored at –80°C, and it was subsequently used for RNA extractions, ELISA and protein gels.

RNA Isolation

RNA was isolated from the same purified virus preparation by sucrose density gradients, phenol/chloroform extractions and a Purescript RNA extraction kit (Gentra Systems, Minneapolis, MN).

Linear log sucrose density gradients were prepared (Brakke and Van Pelt, 1970) and allowed to form overnight at 4°C. An aliquot of the purified virus preparation, containing approximately 3.4 mg/ml, was incubated at room temperature (RT) for 10 min with an equal volume of RNA dissociating solution (200 mM Tris-HCl, 2 mM EDTA, 2% sodium lauryl sulphate, pH 9.0) and 6 µl of protease K (10 mg/ml). The preparation was layered onto the gradients, which were centrifuged in a Beckman SW 41 rotor at 185,000 *g* for 5 h at 15°C. Gradient zones containing RNA were collected using an ISCO UV fractionator (ISCO Inc., Lincoln, NE). The RNA was precipitated by addition of three volumes of 100% ethanol and 0.1 volume of 3 M sodium acetate (pH 5.2), and incubation overnight at -20°C (Sambrook *et al.*, 1989). After centrifugation at 10,000 *g* for 10 min, the RNA pellet was washed in 70% ethanol and vacuum dried. The RNA was resuspended in 50 µl of molecular biology grade water (MBG) and stored at –80 C.

RNA was also extracted from virions by incubating 100 µl purified virus with 100 µl RNA dissociating solution and 200 µl protease K (10 mg/ml) at RT for 10 min. An equal volume of phenol:chloroform:isoamyl alcohol (1:1:24) was added and the mixture was centrifuged at 12,000 *g* in a microfuge for 5 min. The aqueous phase was collected,

an equal volume of chloroform:isoamyl alcohol (1:24) was added, and the mixture was centrifuged at 12,000 *g* for 5 min. The RNA was precipitated by addition of 2.5 volumes of 100% ethanol and 0.5 volume of 7.5 M ammonium acetate, and incubation overnight at -20°C (Sambrook *et al.*, 1989). After centrifugation at 10,000 *g* for 10 min, the RNA pellet was washed in 70% ethanol and vacuum dried. The RNA was ressuspended in 40 µl of molecular biology grade (MBG) water and stored at –80° C.

An RNA extraction procedure using the Purescript RNA extraction kit was performed according to manufacturer's suggestions, by gently mixing 100 µl purified virus with 300 µl of cell lysis buffer for 5 min. The mixture was heated to 65°C for 5 min and cooled to RT. This was followed by addition of 100 µl of protein/DNA precipitating solution and gentle mixing by inversion. The tube was incubated in an ice bath for 5 min and centrifuged at 13,000 *g* for 4 min, Then 400 µl of the supernatant was collected and 300 µl of isopropanol was added to it. The tubes were inverted gently for 50 times to mix, and the RNA was pelleted by centrifugation at 16,000 *g* for 20 min at 4°C. The supernatant was removed, the pellet was washed with 70% ethanol, and vacuum dried. The RNA was resuspended in 40 µl MBG water and stored at –80° C.

Analysis of Extracted RNA

The RNA obtained from all extraction procedures was analyzed by electrophoresis in agarose gels, with and without the addition of formamide to the sample. Formamide was deionized by mixing 5 ml of formamide with 0.5 g of fully regenerated analytical grade mixed-bed resin (AG 501-X8, 20-50 mesh) (Bio-Rad Laboratories, Hercules, CA). The formamide and resin mixture was left on a shaker for 1 h at RT, and then filtered

through 2 X Whatman no.1 filter paper, divided into aliquots and stored at –80°C. Samples for electrophoresis were prepared by adding 4 µl of MBG water and 5 µl deionized formamide, to 1 µl RNA. The samples were then incubated at 50°C for 5 min, allowed to cool on ice, and 1 µl of gel loading buffer (GLB) (prepared with l-diethyl-pyrocarbonate [DEPC] treated water) was added to the sample. Samples were also prepared without formamide, by the addition of 9 µl MBG water and 1 µl GLB to 1 µl of RNA. RNA samples were electrophoresed in 0.9% agarose gels with TAE 1 X buffer prepared using DEPC treated water.

Synthesis of cDNA

Reverse transcription of RNA was performed using Superscript II (Life Technologies, Gaithersburg, MD) according to manufacturers instructions, and a specific primer within the 3' end of the region coding for the cylindrical inclusion protein, 5' GCGAATAACAACGATATG 3' (EH 281). Second strand synthesis was performed by adding to the first strand reaction, second strand reaction buffer, DNA polymerase I, DNA ligase, and RNAse H from *E. coli*, and T_4 DNA polymerase, according to manufacturer's instructions. Approximately 0.3 µl of [^{32}P] dCTP (10 µCi) (3,000 Ci/mmol) (Du Pont NEN, Boston, MA) were added to aliquots of 3 to 5 µl of the first strand, or second strand reactions, and used as a tracer. Size analysis of the radio-labeled products of the reverse transcription and second strand reactions were performed on 0.9% alkaline agarose gels (Lambda Librarian, Stratagene Cloning Systems, La Jolla, CA). The gel was exposed to X-ray film and compared to a 1 kb ladder molecular weight standard (Life Technologies, Gaithersburg, MD). The resulting cDNA products were also analyzed by

electrophoresis in 0.9% agarose gels stained with ethidium bromide, and were used for polymerase chain reaction amplification (PCR) of viral cDNA and for 5' rapid amplification of cDNA ends (5' RACE).

Polymerase Chain Reaction

PCR amplifications of double stranded cDNA were performed using primer EH 281 and random primers (Pharmacia Biotech Inc., Piscataway, NJ). Super Mix Taq polymerase mix and Elongase (Life Technologies, Gaithersburg, MD), were used to obtain clones representing the 5' terminal region of the genome of WLMV. Elongase is an enzyme mixture containing a polymerase and a proof-reading enzyme, and it is designed for amplification of long templates. The resulting PCR products were gel purified by electrophoresis in 1% agarose, and extracted from the agarose using either GEN-Elute agarose spin columns (Supelco, Bellefonte, PA) or the QIAquick Gel Extraction Kit (QIAGEN Inc., Valencia, CA) according to manufacturers' instructions. DNA fragments were ligated by A-overhangs into pGEM-T-Easy (Promega, Madison, WI), or blunt end ligated into Clonesure pGATA (Clonesure, Lincoln Park, NJ). *E. coli* DH5α was transformed with the ligation products and DNA was extracted by alkaline lysis miniprep, and digested with EcoR I (Promega, Madison, WI) to determine insert size and orientation. The largest clones were sequenced using the specific primer and a vector primer. Based on the sequence obtained, a new primer was designed within the 5' end of the region encoding for the CI, 5' GGGTGTTGTGTTGTGAAAGAT 3' (EH 305). The process was repeated once, generating clones within the region encoding for the CI and P3 protein, and a primer within the P3 encoding region 5' CAGTGCCATAAGATTTTACCTTCGC 3' (EH 319).

PCR products obtained from amplification of cDNA using primer EH 319 and random primers were purified using the QIAquick PCR Purification Kit (QIAGEN Inc., Valencia, CA) and ligated into pGEM-T-Easy overnight. The ligation reaction was used to transform *E. coli* DH5α, and clones were selected by blue-white color selection and were screened for size using a quick miniprep procedure. DNA was purified by alkaline lysis miniprep and digested with EcoR I to determine insert size and orientation, and two clones were sequenced using primer EH 319 and vector primers.

Synthesis of a 5' terminus cDNA library

In order to obtain large clones representing the 5' terminal half of the virus, a cDNA library was generated from viral RNA using a specific primer within the cylindrical inclusion encoding region (EH 281), random primers, and the "Time Saver" cDNA synthesis kit (Pharmacia Biotech Inc., Piscataway, NJ) according to manufacturer's instructions. RNA was extracted from purified virions using protease K and RNA dissociating solution, followed by treatment with phenol:chlorofom:isoamyl alcohol (1:1:24) and ethanol precipitation. An aliquot (3 µl) of the first strand reaction was removed and 1 µl of [^{32}P] dCTP (10 µCi) (3,000 Ci/mmol) (Du Pont NEN, Boston, MA) was added. Blunt-ended cDNA fragments were ligated to EcoR I/Not I adapters and inserted into EcoR I sites of Lambda Zap II (Stratagene, La Jolla, CA). The ligated Lambda DNA was packaged using the "Packagene" Lambda DNA Packaging system (Promega, Madison, WI).

To determine the titer of recombinant phage, color selection was used. Aliquots of the packaged library were added to *E. coli* strain XL-1 Blue MRF' grown to mid-log phase and adjusted to $OD_{600} = 0.5$ in 10 mM $MgSO_4$. Cells were mixed with 15 µl 0.5 M

isopropyl - β - thiogalactopyranoside (IPTG) and 50 µl X-galactose (X-gal) (250 mg/ml) in 3 ml melted NZY top agar (0.5% NaCl, 0.2% $MgSO_4$7 H_20, 0.5% yeast extract, 1% casein hydrolysate and 0.8% agar). This mixture was immediately plated onto NZY plates (0.5% NaCl, 0.2% $MgSO_4$.7 H_20, 0.5% yeast extract, 1% casein hydrolysate and 1.5% agar) and incubated overnight at 37°C. Plates were incubated at 4°C for 3 h for color enhancement and plaques were counted.

Immunoscreening of WLMV clones

Amplification of the cDNA library made to the 3' terminal region of WLMV and immunoscreening for clones expressing the CI protein and the helper component protease (HC-Pro) were done according to the "Lambda Zap II Instructional Manual" (Stratagene), as described in Chapter 2. For detection of fusion proteins the following rabbit polyclonal antisera were used: antiserum (AS 1180) to the cylindrical inclusion (CI) of WLMV (obtained from Dr. D. E. Purcifull), and antiserum (AS 1136) to the amorphous inclusion (AI) of papaya ringspot virus type W, isolate W1A (PRSV-W, W1A) (Baker, 1989).

The cDNA library prepared to the 5' terminal region of WLMV was immunoscreened for clones expressing the HC-Pro using AS 1136 and an antiserum prepared to the AI of tobacco vein mottling virus (TVMV) (Thornbury *et al.*, 1985).

Nucleic acid hybridization of WLMV clones

Biotinylated probes were prepared using clones of known sequence and the Non-Radioactive Random Octamer Labeling System (Tropix, Inc., Bedford, MA). Efficiency of biotinylation of each probe was tested according to manufacturer's instructions using dilutions ranging from 100 µg/µl to less than 1 pg/µl. The Southern-Light protocol (Tropix, Inc., Bedford, MA) was used for detection of the chemiluminescent signal.

Aliquots of the recombinant library were plated according to manufacturer's instructions; plates were incubated at 37°C overnight and then at 4°C for 4 h. Sterile nylon Hybond membranes (0.45 μ pore size, Amersham, Inc., Arlington Heights, IL), cut to size, were laid on plates containing well separated plaques for 3 min at RT. The membranes were incubated for 3 min in denaturing solution (0.5 M NaOH, 1.5 M NaCl), two times. After the denaturation step the membranes were equilibrated twice in 2 ml of 1 M TRIS buffer (pH 7.5), for 3 min, and were allowed to air dry.

The dry membranes were placed in a UV crosslinker for 2 min at optimal crosslinking (1,200 x 100 $\mu J/cm^2$), transferred to hybridization bottles and washed with distilled water. The membranes were prehybridized in 10 ml of hybridization buffer (1 mM EDTA, 7% SDS, 0.25 M disodium phosphate, pH 7.2) and incubation overnight in a Hybaid (Vangard Int., Inc., Neptune, NJ) rotating oven at 65°C. For hybridization, 70 μl aliquots of the probe, containing 800 μl of salmon sperm DNA and 11 ml of hybridization buffer were boiled for 10 min and immediately placed on ice. The prehybridization solution was discarded and the prepared probe was added to the hybridization bottles and incubated for 2 h at 65°C. The probe was recovered and frozen at –20°C and reused when needed. Membranes were washed twice in 2 X SSC containing 1% SDS, for 5 min at RT, twice in 0.1 X SSC containing 1% SDS for 15 min at 65°C, and twice in 1 X SSC for 5 min at RT.

Hybridization of the probe to the membranes was followed by chemiluminescent detection of the biotinylated DNA according to manufacturer's instructions. All incubations and washes were performed in a Hybaid rotating oven at RT. The membranes

were washed twice in blocking buffer (0.2% I-Block Reagent, 1 X PBS, 0.5% SDS) for 5 min, once in blocking buffer for 10 min, and incubated for 20 min in Avidix AP diluted 1:8,000 in blocking buffer. The membranes were then washed once in blocking buffer for 5 min, three times in wash buffer (1 X PBS, 0.5% SDS) for 5 min, and twice in assay buffer for 2 min. The bottles were drained, 900 µl of CDPStar was added, and the incubation was allowed to proceed for 5 min with gentle mixing by hand. CDPStar was recovered and reused by adding 400 µl fresh CDPStar. The hybridization bottles were drained and the membranes were placed in a developing folder and exposed to XAR X-ray film (Eastman Kodak, Rochester, NY) for periods of 20 min to 2 h. Films were developed and used as templates for plaque plug collection from plates.

Positive Lambda Zap II clones were plaque purified by immunoscreening three times, or by screening with the biotinylated probe twice. The pBlueScript SK (-) phagemid contained in the vector was excised as described in Chapter 2. For isolation of single *Escherichia coli* SOL R colonies containing the excised phagemids, bacterial mass was streaked onto a new plate, and at least two isolated colonies were amplified for clone analysis.

5' Rapid Amplification of cDNA Ends

The 5'UTR of WLMV was amplified by 5' rapid amplification of cDNA ends (5' RACE) (Schaefer, 1995). Purified RNA was reverse transcribed using Superscript II with a primer designed for amplification of the P1 coding region, 5' AACGCATTTGAAAAACAGCG 3' (EH 325), and the cDNA was polyadenylated using terminal deoxynucleotidyl transferase (Promega, Madison, WI). The polyadenylated cDNA was amplified by two cycles of PCR amplifications using DyNAzyme EXT

(Fynnzymes Oy, Espoo, Finland), an optimized mixture of DNA polymerase and a proofreading enzyme, with the accompanying buffer optimized for amplification of short templates. PCR amplifications were performed using an oligo dT anchor primer, 5' GACCACGCGTATCGATGTCGACTTTTTTTTTTTTTTT(A/G/C) 3', and a PCR anchor primer, 5' GACCACGCGTATCGATGTCGAC 3' (both primers were generously provided by Dr. K. Derrick), and an internal primer. PCR products were gel purified and ligated into pGem-T-Easy, *E. coli* DH5a was transformed with the ligation products and clones were screened for size and orientation using restriction enzymes. The two largest clones were sequenced.

Analysis of WLMV clones

Plasmid clones (pBlueScript SK-) were screened by a rapid miniprep procedure and were prepared for further analysis by alkaline lysis miniprep. Screening for insert sizes and sequence direction was done by PCR using specific internal primers (EH 281, EH 305, EH 319, EH 322), a T3 primer (previously synthesized at the DNA Synthesis Core, University of Florida), and a primer complementary to the T7 sequence of pBlueScript SK- (EH 276). Clones obtained from Lambda Zap II libraries and by PCR, were screened for insert size by digestion with EcoR I and were mapped using restriction enzymes (Nsp V, EcoR I, Hind III, Xho I, Xba I and Sac I). Plasmid DNA was purified for sequencing using the QIAprep Spin Miniprep Kits (QIAGEN Inc., Valencia CA). All sequencing was done by automated sequencing at the DNA Sequencing Core Facility of the Interdisciplinary Center for Biotechnology Research (ICBR) at the University of Florida.

Fragment assembly and analyses of nucleotide and predicted amino acid sequences were done using the Wisconsin Package Version 8.1, (Genetics Computer Group [GCG],

Madison, WI.) and the Wisconsin Package Version 9.0 available at the ICBR Biological Computing Facility, University of Florida. Sequences of potyviruses, listed on Table 3-1, were chosen from the output of BLAST and BLASTP (Altschul *et al.*, 1990) and obtained from GenBank. The sequences were aligned, consensus regions were compared, and neighbor-joining trees were calculated using the neighbor-joining distance matrix method of Saitou and Nei as implemented in CLUSTAL W (1.7) (Thompson *et al.*, 1994). Reliability of the trees was estimated using bootstrap with 1,000 replicates. Trees generated by CLUSTAL W (1.7) were viewed using TreeView 1.2 (Page, 1996). Theoretical molecular size was calculated by Compute pI/Mw. PSORT was used to predict cellular localization of the WLMV proteins and to identify possible nuclear localization signals (NLS). PSORT II also was used to find putative NLS. Tm-Predict was used to predict transmembrane regions of P3 protein. LALIGN was used to calculate the percentage identities between the polyprotein of WLMV and the polyprotein of nine viruses in the *Potyvirus* genus and one in the *Rymovirus* genus (brome streak virus, BStV) (Gotz and Maiss, 1995) listed in Table 3-1.

Results

Virus Purification and RNA extractions

WLMV was easily purified, and virus yield was 14.1 mg/100 mg leaf tissue. A large number of particles was seen in the uranyl acetate stained purified virus preparation examined by electron microscopy, and there was no obvious evidence of heterogeneity in particle size. Extraction of viral RNA using RNA dissociating solution and protease K,

followed by phenol:chloroform:isoamyl alcohol treatment and ethanol precipitation yielded a large amount of undegraded RNA, judged by agarose gel analysis.

Synthesis of cDNA and PCR clones

Synthesis of double-stranded (ds) cDNA from WLMV RNA, using primer EH 281 at the 3' end of the CI encoding region, resulted in a cDNA of approximately 5 kbp, but in very low concentration as indicated by agarose gels. The largest cloned fragments obtained by PCR amplification of WLMV ds cDNA using Super Mix and Elongase, and using one specific primer and random primers, were 0.7 kbp and 1.6 kbp, respectively. PCR products of up to 3 kbp were obtained in low concentrations, judged by agarose gel analysis, but were not cloned successfully into either pGem T-E or into pGATA. Therefore, a Lambda Zap II library was generated for the 5' terminal region of the genome of WLMV. Clones obtained by PCR overlapped with clones obtained from the Lambda Zap II libraries and were used to determine and confirm the sequence of WLMV.

Clones representing the genome of WLMV

After serological screening of the plaques generated from the cDNA library to the 3' end of WLMV with the antiserum prepared to the CI of WLMV (AS 1180), eleven immuno-positive clones were identified and excised. The insert sizes of six selected clones were further analyzed after EcoR I digestion (Table 3-2). Three clones (pCIA-6, pCIB-1 and later pCIB-2) were selected for sequencing with M13 forward and reverse primers. Clone pCIA-6 overlaps with pCPB (Chapter 2) within the replicase coding region. A unique Nsp V restriction enzyme site allowed for the ligation of the two clones, at the 3' end of clone pCIA-6. Clone pCIB-2 overlaps with the 5'end of clone pCIA-6 within the CI coding region and extends into the P3 protein gene (Fig. 3-1). A single Sac I restriction

site found between clones pCIA-6 and pCIB-2 may be useful in ligating these two clones. Clone pCIB-1 is entirely within the CI coding region.

Serological screening with AS 1136, prepared to the amorphous inclusion (AI) of PRSV-W, and with an antiserum prepared to the AI of TVMV, resulted in nonspecific labeling of plaque lifts and was discontinued.

A biotinylated probe prepared using a clone within the 3' terminal region of the CI coding region (p281/RP) was used to identify clones from the two Lambda libraries, and other clones within the CI gene were isolated and excised. However, no clones containing the P3, HC-Pro or P1 genes were identified. The library prepared for the 5' terminal region of WLMV was screened with probes within the regions encoding for the P3, HC-Pro and P1. Clones representing the P3, HC-Pro, P1 and part of the 5' untranslated region (5'UTR) were isolated and are listed in Table 3-2.

The cloned products of 5'RACE were mapped and two clones were sequenced. Clones obtained by 5'RACE overlapped clones obtained from the cDNA library and 57 nucleotides (nt) were added to the sequence of the 5'UTR. The full-length sequence of WLMV from the putative 5' terminal adenosine to the 23 cloned adenosines of the poly (A) tail is presented in Fig. 3-2. Rare nucleotide divergences were found in overlapping segments, but these did not result in changes in the deduced amino acid sequence, indicating a homogeneous population. Selected clones representing the entire genome of WLMV are presented in Fig. 3-1.

Genome organization of WLMV

The sequence analysis results indicate that WLMV is 9,994 nt long, excluding the poly (A) tail. The single open reading frame (ORF) is 9,657 nt long, beginning at position

95 and ending at the 'UAA' termination codon at position 9,749-9,751, and it encodes a polyprotein of 3,218 amino acids (aa). The number of nucleotides in the 5'UTR and in the 3'UTR, as well as the numbers of amino acids in the viral proteins, of WLMV are compared to those of other potyviruses in Table 3-3, showing size variability. The ORF is preceded by a 5'UTR of 94 nt with an adenosine at the 5' terminus, and is followed by a 3'UTR of 243 nt, and a poly (A) tail. An alignment of sequences of WLMV and other potyviruses showing the consensus sequences for the potybox and boxes a and b, is shown in Fig. 3.3. The sequence at the 5' end of WLMV shows six nt between the 5' terminal adenosine and box a, and 19 nt between box b and the initiation codon (AUG). The initiation codon is within the sequence context ACAAUGGCAA, similar to that proposed by Lutcke *et al.*, (1987) for initiation of translation in plant RNA.

The first protein at the amino-terminal end of the potyviral polyprotein is the P1 protein, which catalyzes proteolysis between itself and the helper component protease (HC-Pro). The P1 protein of WLMV is 433 aa long with a calculated size of 50 kDa. The consensus motif encompassing the active site serine residue found in serine type proteases GXSG is strictly conserved in potyvirus P1 proteins (Verchot *et al.*, 1991), and has the sequence GNSG in WLMV. The motif FILRG believed to be important for protein processing (Verchot *et al.*, 1991), is strictly conserved in WLMV. The cleavage site between P1 and HC-Pro has the sequence YKLG/G (Fig. 3-4). Two possible nuclear localization signals (NLS) were found using PSORT, HKKR and KKRR. PSORT II identified several possible NLS in P1: HKKRR, PGFQRRK, PVKVRRTR, and PHQDRKR.

The second protein of the WLMV polyprotein is the helper component-protease (HC-Pro), with 460 aa and calculated size of 52.6 kDa. An alignment of conserved regions believed to be important for the helper component function of HC-Pro, is shown for WLMV and nine other potyviruses in Fig. 3-4. The amino-terminal domain of HC-Pro is involved in mediating aphid transmission and has a strongly conserved, cysteine-rich region that is postulated to form a zinc finger-like motif with structural or functional roles, with the following consensus sequence: $HXCX_8CX_{18}CX_2C$ (Atreya *et al.*, 1992). The PTK and KITC motifs are shown in Fig. 3-4. The PTK motif has been shown to bind to the virion (Peng *et al.*, 1998). The KITC motif is important for the interaction between HC-Pro and aphid stylets (Huet *et al.*, 1994), and is found in WLMV with the sequence RITC. The carboxy terminal region of HC-Pro is a protease that autocatalytically cleaves itself at the YKIG/G cleavage site between the HC-Pro and P3 proteins. The prediction of cellular localization using PSORT indicates accumulation of the HC-Pro in the cytoplasm (65%), and no nuclear localization signals were identified. However, a search using PSORT II identified a possible nuclear localization signal at PNGKRKL beginning at aa 244.

The P3 protein is 346 aa long with a calculated size of 39.8 kDa. The P3 is believed to be a transmembrane protein and PSORT prediction localizes the P3 of WLMV to the plasma membrane, with a transmembrane region from aa 254 to aa 270. Transmembrane helices were also identified by TM-predict.

The major viral protease is the genome-linked protein protease (VPg-Pro), which cleaves the polyprotein at six cleavage sites defined by the consensus sequence VXXQ/(A, G, S, V). The VPg-Pro cleavage sites of WLMV are: VHFQ/A between the P3 and 6K1

proteins; VHFQ/S between the 6K1 and the CI proteins; VHFQ/S between the CI and the 6K2 proteins; VHFQ/G between 6K2 protein and the VPg-Pro; VHEQ/S between the VPg-Pro and the replicase; and VHFG/S between the replicase and the coat protein (Fig. 3-2). A suboptimal cleavage site, VMFE/S, within the amino terminus of VPg-Pro, cleaves off the genome-linked protein (VPg) that is found covalently linked to the 5'terminus of the viral RNA. The proposed consensus for the amino acid residues surrounding the tyrosine residue XFXNMYGX(S,D)PXX(S,N)XX, is found in WLMV with the sequence: KFIHMYGVDPSEY (Murphy *et al.*, 1991).

The VPg-Pro of WLMV is 434 aa long with a calculated size of 49.2 kDa, and accumulates in the nucleus with the replicase forming the nuclear inclusion a (NIa) and nuclear inclusion b (NIb) (Petersen *et al.*, 1996). A possible bipartite nuclear localization signal beginning at aa 4 with the following sequence KRQRQKLRFRDTRDRRL, was identified in VPg-Pro using PSORT. Fig. 3-5 shows an alignment of the VPg-Pro of WLMV and of nine other potyviruses, indicating the cleavage sites between the 6K2 protein and the VPg-Pro, putative NLS and the consensus around the tyrosine residue that links the VPg to the viral RNA.

The 6K1 protein was found to be 52 aa residues long with a calculated size of 5.6 kDa. The 6K2 protein was found to be 54 residues long with a calculated size of 5.9 kDa. PSORT prediction identified a transmembrane segment from aa 7 to aa 23 in 6K1 and primarily associated with the plasma membrane and secondarily to the membranes of the endoplasmic reticulum. The 6K2 protein was also predicted to be localized to the endoplasmic reticulum as a transmembrane protein with a transmembrane segment from aa 24 to aa 40.

The cylindrical inclusion protein (CI) of WLMV is made up of 636 aa residues with a calculated size of 71.6 kDa. A highly conserved motif found in all potyviruses, GAVGSGKST, was identified in the CI protein of WLMV. This motif is agreement with the nucleoside triphosphate binding motif (A,G)XXXXGK(S,T) found in ATPases (Lain *et al.*, 1991). Other consensus motifs present in CI proteins are the DEXH domain with the sequence DECH also involved in NTPase activity. The DEXH motif is believed to interact with the consensus motif QXXGRXXR' which confers RNA binding activity to the helicase-like proteins, and is within a highly conserved region in potyviruses (Fernandez and Garcia, 1996) with the sequence GERIQRLGRVGR in WLMV.

The replicase of WLMV is 520 aa long with a calculated size of 59.9 kDa. A possible nuclear localization motif was found at aa 303 with the sequence KKHK. The replicase is the most conserved gene product among potyviruses, with percentages of amino acid sequence identities ranging from 61 to 64%. Regions of high conservation were found in the sequence of the replicase. The consensus motif (T,S)GXXXTXXXNS(T,S) (21 to 38 aa) GDD found in RNA-dependent RNA polymerases was found 166 amino acids upstream of the first serine of the coat protein with the sequence SGQPSTVVDNSL (29 aa) GDD.

The coat protein is 284 aa long and has a calculated size of 32.6 kDa. A DAG triplet, important for aphid transmission, was found at position 25 downstream of the initial serine on the highly variable amino terminal domain of the coat protein. Within the core region, three amino acids highly conserved in filamentous and rod shaped viruses were found, a serine residue at position 152, an arginine at position 184, and an aspartate at position 218. The latter two residues have been postulated to form a salt bridge

important for conservation of the capsid protein structure in all filamentous viruses (Dolja *et al.*, 1991).

The 3' untranslated region (3'UTR) is 243 nt long and was predicted to fold into a stable secondary structure using an RNA folding program MFOLD (Zuker, 1989).

A CLUSTAL W (1.7) tree was generated from alignments of the full-length polyprotein sequences of WLMV and of nine other potyviruses, and is shown in Fig. 3-6. High bootstrap support was found for a branch placing WLMV in a clade with peanut stripe virus (PStV) and zucchini yellow mosaic virus (ZYMV).

Percentage of identity between the WLMV polyprotein and those of nine other potyviruses ranged from 46.4% to 48.4%, and between WLMV and BStV was 27.9%.

Discussion

Clones representing the entire genome of watermelon leaf mottle virus (WLMV) were obtained by preparing two cDNA libraries in Lambda Zap II, and by PCR amplification of selected regions of the viral genome. Clones representing regions of the genome encoding part of the P3 protein, the two 6K proteins, the CI protein, VPg-Pro, the replicase, the coat protein, and the 3'UTR and part of the poly (A) tail, were obtained from the cDNA library generated from WLMV RNA using oligo dT and random primers. Clones representing the coding regions for the CI protein, the P3 protein, the HC-Pro, the P1 protein, and part of the 5'UTR, were obtained from the cDNA library generated from WLMV using a primer within the CI gene and random primers. The remainder of the 5'UTR was obtained by 5'RACE.

Large clones representing the entire genome of WLMV have been isolated, and these may be ligated to make a full-length clone. Sequence analyses point to some regions that may be of interest for further studies. Predictions of the cellular localization of proteins and identification of cellular sorting signals are not conclusive evidence of localization, but may be good indicators of regions in the viral genome that may be of interest for studies of protein function. Experimental evidence is needed in order to confirm the initial predictions.

Entry of proteins into the nucleus is a complex process involving recognition and binding of a nuclear localization signal (NLS) to the nuclear pore and an energy dependent transport into the nucleus. Possible NLS were found in the P1 protein, the HC-Pro, VPg-Pro and the replicase. Although the identified NLS have not been experimentally proven, nuclear inclusions made up of NIa and NIb are found in the nuclei of cells infected with WLMV (Petersen *et al.*, 1996). Localization of the VPg-Pro and the replicase proteins of tobacco etch virus (TEV) to the nucleus is associated to NLS (Restreppo *et al.*, 1990; Carrington *et al.*,1991) and has been reported to be sensitive to sequence alterations (Li *et al.*, 1997).

Nuclear localization signals have not been previously reported in the sequences of the HC-Pro or P1 proteins. However, immunogold-labeling experiments using antiserum to the AI of PRSV-W (Petersen *et al.*, 1996), have identified an unusual inclusion in the nuclei of squash and watermelon infected with WLMV. Furthermore, the nuclear inclusions of PStV and beet mosaic virus (BtMV) were labeled with antiserum prepared to the expressed HC-Pro, HC-VPg and the replicase of PPV (Riedel *et al.*, 1998). PSORT predictions of the P1 proteins of the other viruses used in the alignments indicated putative

NLS not only in WLMV, but also in PRSV, PStV, and TEV. Experimental evidence is needed to confirm if the putative NLS are effective in localizing the P1 and the HC-Pro to the nucleus. This information may be of interest because the P1/HC-Pro sequence may function as a general suppressor of virus-induced gene silencing (Anadalakshmi *et al.*, 1998; Kasschau and Carrington, 1998).

Table 3-1. List of abbreviations, accession numbers and names of the viruses used in the sequence alignments for comparison with watermelon leaf mottle virus (WLMV). Sequences were obtained from GenBank after BLAST search with full-length sequence of WLMV.

Abbreviation	Accession Number	Virus name
PPV-RA	P17767	Plum pox virus Rankovic
PRSV-Ha	X97251	Papaya ringspot virus - Hawaii
PSbMV	X89997	Pea seed borne mosaic virus
PStV	U34972	Peanut stripe virus
TEV	P04517	Tobacco etch virus
TuMV	D83184	Turnip mosaic virus
TVMV	X04083	Tobacco vein mottling virus
ZYMV-CA	L31350	Zucchini yellow mosaic virus - California
ZYMV-RU	L29569	Zucchini yellow mosaic virus - Reunion Island

Table 3-2. Watermelon leaf mottle virus (WLMV) cDNA clones obtained from Lambda Zap II libraries. Size approximation of clones was done by digestion with EcoR I, and electrophoresis in 0.9 % agarose gels.

Probe Designation	Clone Designation	Approximate Size (kb)
AS 1180 [a]	pCIA-2	0.9
	pCIA-3	1.0
	pCIA-4	1.2
	pCIA-6	2.8
	pCIB-1	1.6
	pCIB-2	3.2
p281/RP [b]	p1D1a	4.5
	p1D7	1.4
p305/RP [b]	p1-1a	3.0
	p1-2a	1.2
	p2-1a	1.4
	p3-4a	2.2
	p3-5a	1.9
	p3-7a	1.9
p319/RP4 [b]	p3-3-2	2.3
	p3-2-1	2.2
p321 [b]	p1-1-1	2.4
	p1-2-1	0.8
	p1-3-1	0.8
	p1-4-1	0.8
	p1-5-1	2.4
	p1-7-1	0.9
	p1-8-1	0.7

[a] Clones identified by immunoscreening with antiserum to the cylindrical inclusion protein of WLMV (AS 1180).
[b] Clones identified by screening with biotinylated cDNA probes.

Table 3-3. Comparison of watermelon leaf mottle virus (WLMV) and other potyviruses, showing the number of nucleotides (nt) in the 5' untranslated region (5'UTR) and in the 3' untranslated region (3'UTR), and the number of amino acids (aa) in each protein.

Region of the Genome	WLMV[a]	PRSV[b]	ZYMV-CA[c]	ZYMV-RU[d]
5'UTR (nt)	94	85	140	144
P1 (aa)	433	547	310	314
HC-Protease (aa)	460	459	456	455
P3 (aa)	346	346	346	346
6K1 (aa)	52	52	52	52
CI (aa)	636	635	634	634
6K2 (aa)	53	57	53	53
VPg-Protease (aa)	434	427	433	433
Replicase (aa)	519	537	517	517
Coat Protein (aa)	283	287	279	279
3'UTR (nt)	243	206	214	245
Polyprotein (aa)	3,216	3,347	3,080	3,083
RNA (nt)	9,994	10,326	9,593	9,642

[a] Watermelon leaf mottle virus (WLMV).
[b] Papaya ringspot virus (PRSV).
[c] California strain of zucchini yellow mosaic virus California (ZYMV-CA).
[d] Reunion Island strain of zucchini yellow mosaic virus (ZYMV-RU).

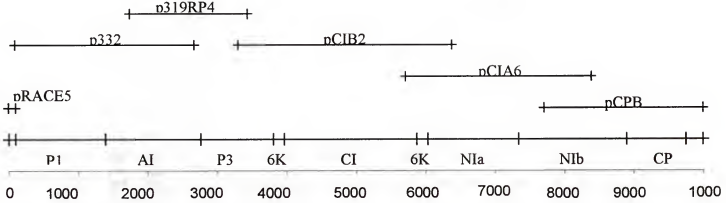

Fig. 3-1. Map of selected cDNA clones representing the entire genome of watermelon leaf mottle virus (WLMV).

Fig. 3-2. Full length sequence of watermelon mosaic virus (WLMV). The predicted amino acid sequence of the open reading frame for the putative polyprotein is shown. Underlined nucleotide sequences are "box a" and "box b", and the initiation of translation consensus sequence. Underlined amino acid sequences correspond to the protease cleavage sites.

```
1     GAAATCTCAACACAACATAACGAAAACGATCTCATTCACCTTCAAACACAAACGCTTTCA 60

61    ATCACATTCAAAGCAATTCTTCTCACATATTACAATGGCAACACTGGTTTTTGGATCTTT 120
                                    M  A  T  L  V  F  G  S  F

121   CGTCACTCAAGTCACTGCACCTGCAACCAAAATTAAGACAATGAAAAAGACACGCGTGAT 180
       V  T  Q  V  T  A  P  A  T  K  I  K  T  M  K  K  T  R  V  I

181   CATGAAACCGCACTTATCTGTCATAACAATACCTGAAGTCATGGGAAACGCGTATACTAA 240
       M  K  P  H  L  S  V  I  T  I  P  E  V  M  G  N  A  Y  T  K

241   GTGCTTCTGTGGCACTATTTGCCCGACCAAATCTTTTTGGCGCAAACACGTCGCATCGAA 300
       C  F  C  G  T  I  C  P  T  K  S  F  W  R  K  H  V  A  S  K

301   GTGCGACAAGGATTGCACCGATGCGGAACTTGCGCAAGTCGGAGCAAAACGCAAAACCTA 360
       C  D  K  D  C  T  D  A  E  L  A  Q  V  G  A  K  R  K  T  Y

361   TGAGCATGAGACGATCCTCGTACCTGTCGAGGAAGAGTATGAAGATGTAATATATGATAG 420
       E  H  E  T  I  L  V  P  V  E  E  E  Y  E  D  V  I  Y  D  S

421   TGTACCTATGGTGTTTAACTTTTCTGATGAGCCAACATATGTAACAAAACCAATCATTGT 480
       V  P  M  V  F  N  F  S  D  E  P  T  Y  V  T  K  P  I  I  V

481   GGAATCACCCAAGCAACGAGATGTCACAACACACGATTTAAGGGTGGATACCAGTTGTAA 540
       E  S  P  K  Q  R  D  V  T  T  H  D  L  R  V  D  T  S  C  K

541   GGACCTGTCACAAGACATTGTGACCTCTCTCTACCAGGACCTGCGTAAACCATTATGTGT 600
       D  L  S  Q  D  I  V  T  S  L  Y  Q  D  L  R  K  P  L  C  V

601   TACACCCAGTAACACATTCGAAGCTTATCGCTGTTTTTCAAATGCGTTTGCAAAACAGCA 660
       T  P  S  N  T  F  E  A  Y  R  C  F  S  N  A  F  A  K  Q  H

661   TGTTTCGGAGATGGTTAAGCTTGATAAATTCGTAGAGGAACGTCTGGCGCAACAAAATTC 720
       V  S  E  M  V  K  L  D  K  F  V  E  E  R  L  A  Q  Q  N  S

721   CAACACAGTCCTGAAGTGTAAGAAAATCAATGGTAGGTTCTGCATTACGCGTATCAAGCC 780
       N  T  V  L  K  C  K  K  I  N  G  R  F  C  I  T  R  I  K  P

781   ATCAGAGCTACCCGGCTTTCAAAGGCGGAAATGGAAAAGAGAGCAAACTGACATAGCAGA 840
       S  E  L  P  G  F  Q  R  R  K  W  K  R  E  Q  T  D  I  A  E

841   GAACATCATCACAAAGCTTTCAGACGTTGAACGATATGAACAGTTTTGCATGGAGGATAG 900
       N  I  I  T  K  L  S  D  V  E  R  Y  E  Q  F  C  M  E  D  S

901   TATAGCTCAGGGAATAAAGAGACCGGGGAAGATACATCCAAAGGTACGGAGAACACGAGC 960
       I  A  Q  G  I  K  R  P  G  K  I  H  P  K  V  R  R  T  R  A

961   CAAACCTGAACGCTCGAACGCAAGCGTGGCCAACTTGTTTCGTTTTGTTGTTAACATTAG 1020
       K  P  E  R  S  N  A  S  V  A  N  L  F  R  F  V  V  N  I  S

1021  CAAACAAGATGGTAAGCAAATCGAGATAATCCATAAGAAACGGAGGATAATTACATTCAG 1080
       K  Q  D  G  K  Q  I  E  I  I  H  K  K  R  R  I  I  T  F  S

1081  TCGTGTTCGGAATAACAATTGTGCACATGTTCGCCTCCCTCACCAGGACCGCAAGCGTTT 1140
       R  V  R  N  N  N  C  A  H  V  R  L  P  H  Q  D  R  K  R  L
```

```
1141  AAAGAGGGAAATAACACTGTCAAGGGATGAAGAACCATACTTTCACATTATGGCGGAGAA 1200
        K  R  E  I  T  L  S  R  D  E  E  P  Y  F  H  I  M  A  E  K

1201  GAATACACGTGGAACACTTAATTACGATGATGTGACGTATGGAAATAGCGGTTTAATATG 1260
        N  T  R  G  T  L  N  Y  D  D  V  T  Y  G  N  S  G  L  I  W

1261  GAAACTTGAAAAAGTAGCAGGGCCAAATCCAGGATACAAAGACATTTTCATATTGCGTGG 1320
        K  L  E  K  V  A  G  P  N  P  G  Y  K  D  I  F  I  L  R  G

1321  AGCATTGAAGAGAGGTCTTGTTTGCGCACAACGGAAAATGAGTTACCAGACAATAAAACT 1380
        A  L  K  R  G  L  V  C  A  Q  R  K  M  S  Y  Q  T  I  K  L

1381  TATTGCCCAGTACTCTAGTAATCCGGGACCTGAATTCTGGGAGGGATTCAATCGTGTGTT 1440
        I  A  Q  Y  /S  S  N  P  G  P  E  F  W  E  G  F  N  R  V  F

1441  TATTAATAGACGAGAAATTGCATTGGATGGACACCAATGCAGGTCACATTACTCAGTCAC 1500
        I  N  R  R  E  I  A  L  D  G  H  Q  C  R  S  H  Y  S  V  T

1501  AAAGTGTGGAGAGATAGCAGCACTTCTTGGCCTATTATTATTACCATCTCATCGTATCAC 1560
        K  C  G  E  I  A  A  L  L  G  L  L  L  L  P  S  H  R  I  T

1561  ATGTGATAAATGTGCACAACAATTTAGTCACATGACATTCTTGGATATAGCGAAGAACAT 1620
        C  D  K  C  A  Q  Q  F  S  H  M  T  F  L  D  I  A  K  N  I

1621  TCGTCAGAAAGCACCGAATCTCCTTGGGATCATACAGAGCAAATACCCAGAATTCGAACA 1680
        R  Q  K  A  P  N  L  L  G  I  I  Q  S  K  Y  P  E  F  E  H

1681  TGTCGCACACTTTCTCAAACACCTTATTGAGATTGAGAACCCAGTGAACGAAAATCCTGA 1740
        V  A  H  F  L  K  H  L  I  E  I  E  N  P  V  N  E  N  P  E

1741  GGATCTAATTCAGATAGCCCAAATTGTTGAACATCGCACCGAGGCCCCGTTCAAATTCAT 1800
        D  L  I  Q  I  A  Q  I  V  E  H  R  T  E  A  P  F  K  F  I

1801  TAAACCAATAAATGAAGCTCTAATACGTAGCACACGAATGACATCGAAAGACATTACATA 1860
        K  P  I  N  E  A  L  I  R  S  T  R  M  T  S  K  D  I  T  Y

1861  TGCTACAAGAAATCTCCTCGAGTTAACGCAATATCTGAAGAACCGAGAAGATAAAATTAA 1920
        A  T  R  N  L  L  E  L  T  Q  Y  L  K  N  R  E  D  K  I  K

1921  AGCAGGATCAGTTGAGAATTTCCGTAACAAAGCAACATCAAAGGCAAGGATGAATCCAAA 1980
        A  G  S  V  E  N  F  R  N  K  A  T  S  K  A  R  M  N  P  N

1981  CCTTTTATGTGACAACCAATTGGACAAAAACGGCAATTTTCTTTGGGGAACACGTGGGTA 2040
        L  L  C  D  N  Q  L  D  K  N  G  N  F  L  W  G  T  R  G  Y

2041  TCACGCAAAGAGAATCTTTACGAATTTCTTCAAAGTGATAGATTCACGTGAACAATATGA 2100
        H  A  K  R  I  F  T  N  F  F  K  V  I  D  S  R  E  Q  Y  E

2101  GGAGTTCACATTGCGAAATGGACCAAACGGTAAGAGAAAACTTGCGATTGGAAACCTCAT 2160
        E  F  T  L  R  N  G  P  N  G  K  R  K  L  A  I  G  N  L  I

2161  TTTACCTGTCGATTTGAAAGAATTACGGAAACATATGGAAGGGGAATCGATTAATCGCAT 2220
        L  P  V  D  L  K  E  L  R  K  H  M  E  G  E  S  I  N  R  I

2221  ACCAGTATCTCAGGCGTGTATTGTTACAAAGCACAATCAATACTTACATCCCTGTTGTTG 2280
        P  V  S  Q  A  C  I  V  T  K  H  N  Q  Y  L  H  P  C  C  C
```

```
2281  TGTGACACATGATGACGGGAGCGTGCTCTATAGTGATCTCATCATGCCAACGAAGAACCA 2340
        V  T  H  D  D  G  S  V  L  Y  S  D  L  I  M  P  T  K  N  H

2341  TCTTGTGATTGGAAATACAGGTGATTCCAAATATGTTGACCTACCACAATTCGATAGTAT 2400
        L  V  I  G  N  T  G  D  S  K  Y  V  D  L  P  Q  F  D  S  I

2401  TCGCATGTACATAGCCGTTGATGGTTATTGTTATCTCAATATCTTCTTGGCAATGCTAGT 2460
        R  M  Y  I  A  V  D  G  Y  C  Y  L  N  I  F  L  A  M  L  V

2461  TAATGTCAATGAACCTGACGCCAAATCCTTTACCAAGATGATTCGAGACAAGCTGATTCC 2520
        N  V  N  E  P  D  A  K  S  F  T  K  M  I  R  D  K  L  I  P

2521  AGAGCTTGGTAAATGGCCAACTGTTCGAGAATTAGCTGCAGCATGTTATTTCCTAGCTGT 2580
        E  L  G  K  W  P  T  V  R  E  L  A  A  A  C  Y  F  L  A  V

2581  CATCTATCCTGAAGTCCGATCTGCAGAGTTACCAAGAATATTGGTTGATCATGACGCTAA 2640
        I  Y  P  E  V  R  S  A  E  L  P  R  I  L  V  D  H  D  A  K

2641  AACGATGCATGTTATTGACTCGTTTGGATCACTACATACCGGATATCACATTCTCAAAGC 2700
        T  M  H  V  I  D  S  F  G  S  L  H  T  G  Y  H  I  L  K  A

2701  TGGCACGGTGAATCAGTTGATTCAATTCAATAGCGATGTGATTATGGGTGAAATGATTGA 2760
        G  T  V  N  Q  L  I  Q  F  N  S  D  V  I  M  G  E  M  I  D

2761  TTATAAAATTGGCGGAACAAGCAAGCCAGAGAGTACAAATTTCGTAGGTAGATTGATCAA 2820
        Y  K  I  G /G  T  S  K  P  E  S  T  N  F  V  G  R  L  I  K

2821  GTCCATTTACAAACCGAAACTCATGGAAGACGTGATAAAAGAGGAACCATACATCTTGTT 2880
        S  I  Y  K  P  K  L  M  E  D  V  I  K  E  E  P  Y  I  L  F

2881  TCATGCACTCACATCACCATCAATCCTCGACGCTCTACACAGGTGTTCACATATAGAGCA 2940
        H  A  L  T  S  P  S  I  L  D  A  L  H  R  C  S  H  I  E  H

2941  CGCAATCAATTTCTGGATTTCAGACAATATCACACTCACGCAGGCGTTAGTTCGTTTATC 3000
        A  I  N  F  W  I  S  D  N  I  T  L  T  Q  A  L  V  R  L  S

3001  ACAACTTGCAAGAAAGGTTCGATATGTCCATGTTCTATCAGCGCAGATGCACATGTTAGA 3060
        Q  L  A  R  K  V  R  Y  V  H  V  L  S  A  Q  M  H  M  L  D

3061  TACAACGGCAGCATCACTATACTCGGAATTATTCCAAGGTTTTGTGATGCAGGATTCTTA 3120
        T  T  A  A  S  L  Y  S  E  L  F  Q  G  F  V  M  Q  D  S  Y

3121  CATGCCCACAATGATGATGTTAAAATTTGTCCAGGACAAGCTCGTCACCAATGAGAAATT 3180
        M  P  T  M  M  M  L  K  F  V  Q  D  K  L  V  T  N  E  K  L

3181  AATAGAATCTGGCTTTGCAACATACGACAGAGAAATTTTGCAAGCACAAGAGAAAAAAGT 3240
        I  E  S  G  F  A  T  Y  D  R  E  I  L  Q  A  Q  E  K  K  V

3241  TCAAAGCGACTTAGCAAATGCATGGTCAGAATTGCGATTGTCGGAAAAGTGTTATTGGCA 3300
        Q  S  D  L  A  N  A  W  S  E  L  R  L  S  E  K  C  Y  W  H

3301  TTGTGTGCATGCAAAGCATACCATTTCACAGGGTGGGCATTACGTAAAGGATGGCATGCC 3360
        C  V  H  A  K  H  T  I  S  Q  G  G  H  Y  V  K  D  G  M  P

3361  AAATTTGAGAGAAATAAAAAGAGCATTTTGCGGAAATGTAATCTCATTGGTGAAACATTC 3420
        N  L  R  E  I  K  R  A  F  C  G  N  V  I  S  L  V  K  H  S
```

```
3421  AGCTAAGAGTGCACATGCGAAGGTAAAATCTTATGGCACTGCTGCACATGCATTTTGTTT 3480
        A  K  S  A  H  A  K  V  K  S  Y  G  T  A  A  H  A  F  C  F

3481  TAGCAAAGTAGTTAATACGTCTTTTAGACTTGTACAATATTGTTTTTCAGATATTTTACG 3540
        S  K  V  V  N  T  S  F  R  L  V  Q  Y  C  F  S  D  I  L  R

3541  TTGTATTAATGTTCTAGTAGTTATGAGTGTTCTTTCTACAATAGCAGTAAATTTAGCGAA 3600
        C  I  N  V  L  V  V  M  S  V  L  S  T  I  A  V  N  L  A  N

3601  CATTGCACATGATAGGAACATAAAAAGAGCAAAGATAAAAGAAATTGAGGCTGAAGGTCG 3660
        I  A  H  D  R  N  I  K  R  A  K  I  K  E  I  E  A  E  G  R

3661  CTTCGATCGATTGTGTACTATTTACAATGCTTATGTTGAGAAGCACGATGAACGACCTAC 3720
        F  D  R  L  C  T  I  Y  N  A  Y  V  E  K  H  D  E  R  P  T

3721  ATTTGATGAATTCATTGATCATTTAAAATACGTTTCACCAGAAGATGTACAGTACCTACA 3780
        F  D  E  F  I  D  H  L  K  Y  V  S  P  E  D  V  Q  Y  L  H

3781  TGAAATCATCAAGAATGATGTACATTTTCAAGCGAAGGGTCCGACAGAAGTGCAATTGGA 3840
        E  I  I  K  N  D  V  H  F  Q  /A  K  G  P  T  E  V  Q  L  E

3841  AAAGATCATGGGTGTACTCATAGTTTTAGCTATGATATTCGACGCAAATAAAAGTGACCC 3900
        K  I  M  G  V  L  I  V  L  A  M  I  F  D  A  N  K  S  D  P

3901  TATAGCAGCCGTTCTAGGTAAGATCAAAAACATCGCCAACACTATGAATCAGAATGTGCA 3960
        I  A  A  V  L  G  K  I  K  N  I  A  N  T  M  N  Q  N  V  H

3961  CTTTCAAAGTCTTGATGAAATTTGTGACATAACTGATGAGAAACAGCTCACTGTCGATTT 4020
        F  Q  /S  L  D  E  I  C  D  I  T  D  E  K  Q  L  T  V  D  F

4021  CGAGTTAAACACGGATACGCATGAACATGTTTCATTAACTGACATAACATTTGCGAAGTG 4080
        E  L  N  T  D  T  H  E  H  V  S  L  T  D  I  T  F  A  K  W

4081  GTGGGAAACTCAACTCGTTGAAAACCGAACAATACCACATTATCGAATTGGTGGTACATT 4140
        W  E  T  Q  L  V  E  N  R  T  I  P  H  Y  R  I  G  G  T  F

4141  CATTGAGTTCACTCGTGCAAATGTCACACAAACTGTCACAACGATAATGCACACTGACGA 4200
        I  E  F  T  R  A  N  V  T  Q  T  V  T  T  I  M  H  T  D  D

4201  TTCGAGCGAATATATGATCAGAGGGGCTGTTGGGTCAGGAAAATCCACGGGTTTGCCATA 4260
        S  S  E  Y  M  I  R  G  A  V  G  S  G  K  S  T  G  L  P  Y

4261  CCACTTATCGTTAAAAGGACATGTTCTTTTGATCGAACCAACGCGACCACTGGCTGAAAA 4320
        H  L  S  L  K  G  H  V  L  L  I  E  P  T  R  P  L  A  E  N

4321  TGTTACACGGCAATTACGAGGAGAACCATTTCACGTCGCACCAACATTGCGCATGCGTGG 4380
        V  T  R  Q  L  R  G  E  P  F  H  V  A  P  T  L  R  M  R  G

4381  CATGTCAAGTTTTGGACATTCACCCATCACCATTATGACATCTGGTTATGCGTTACACCA 4440
        M  S  S  F  G  H  S  P  I  T  I  M  T  S  G  Y  A  L  H  H

4441  CTTTGCCAACAACGTTCAGTCTCTGATTGAATATAAAGTTGTAATGTTTGACGAGTGTCA 4500
        F  A  N  N  V  Q  S  L  I  E  Y  K  V  V  M  F  D  E  C  H

4501  TGTTATGGACGCACAAGCTATGGCATTCTATAGTTTACTGAAAGAACACAATTACAAAGG 4560
        V  M  D  A  Q  A  M  A  F  Y  S  L  L  K  E  H  N  Y  K  G
```

```
4561  GAAGATATTGAAAGTCTCCGCCACTCCACCTGGAAAAGAAGCATCTTTCACAACACAACA 4620
        K  I  L  K  V  S  A  T  P  P  G  K  E  A  S  F  T  T  Q  H

4621  CCCTGTTTCAATACACACAGAGGACCAATTATCATTCGATCAATTTGTTCAAGGTCAAGG 4680
        P  V  S  I  H  T  E  D  Q  L  S  F  D  Q  F  V  Q  G  Q  G

4681  AACTGGATGTAATGCTGATGTCATTTCCAAAGGCAGTAATATTCTGGTGTATGTCGCCAG 4740
        T  G  C  N  A  D  V  I  S  K  G  S  N  I  L  V  Y  V  A  S

4741  TTATAATGAGGTTGACAGCTTGAGTAAAATGTTAATGACGAAGGGTTACGAAGTAACGAA 4800
        Y  N  E  V  D  S  L  S  K  M  L  M  T  K  G  Y  E  V  T  K

4801  GGTGGATGGGCGTACGATGAAAAATGGAAGCACAACTATTGAAACACATGGTACAACAGC 4860
        V  D  G  R  T  M  K  N  G  S  T  T  I  E  T  H  G  T  T  A

4861  CAAACCACATTTTATTGTAGCCACCAATATAATCGAGAATGGAGTCACCATTGACATTGA 4920
          K  P  H  F  I  V  A  T  N  I  I  E  N  G  V  T  I  D  I  D

4921  TTGTGTGGTTGACTTTGGGTTGAAAGTAGTTCCACAACTTGATCTTGATAATAGAATGGT 4980
        C  V  V  D  F  G  L  K  V  V  P  Q  L  D  L  D  N  R  M  V

4981  CCAGTACCGAAAAATACCCGTCAGTTTTGGTGAACGAATACAGCGACTTGGACGTGTTGG 5040
        Q  Y  R  K  I  P  V  S  F  G  E  R  I  Q  R  L  G  R  V  G

5041  ACGTTTCAAACCTGGACACGCACTTCGCATTGGACACACAGAGAAAGGAATAATGGAGAT 5100
        R  F  K  P  G  H  A  L  R  I  G  H  T  E  K  G  I  M  E  I

5101  ACCAACCATGATCGCAACAGAGGCTGCCTTTTTATGTTTTGCATATGGTCTCCCCGTCAT 5160
        P  T  M  I  A  T  E  A  A  F  L  C  F  A  Y  G  L  P  V  M

5161 GACACATAACGTCTCAAATAGCGTAATGGCGAACTGCACAGTGAAACAAGCACGTACCAT 5220
       T  H  N  V  S  N  S  V  M  A  N  C  T  V  K  Q  A  R  T  M

5221  GATGCACTTTGAGCTTTCACCTTTTATAACGTATGAGCTTGTAGATGCTCGTGGAGCAAT 5280
        M  H  F  E  L  S  P  F  I  T  Y  E  L  V  D  A  R  G  A  M

5281  GCACCCTGAAGTGCACCGTGTTTTGAAGCCATTTAAATTGCGTGATTCCGAGATTGAACT 5340
        H  P  E  V  H  R  V  L  K  P  F  K  L  R  D  S  E  I  E  L

5341  CAAAGCCAGTGCCCTTCCAAGCAGTGCCATTTATAGATGGCTAACTATTCAGGATTACGC 5400
        K  A  S  A  L  P  S  S  A  I  Y  R  W  L  T  I  Q  D  Y  A

5401  GAAGATGGGTTGTTATATCGACATAGCCATGGAGACGAAGATTCCTTTCTACATTAAAGG 5460
        K  M  G  C  Y  I  D  I  A  M  E  T  K  I  P  F  Y  I  K  G

5461  CATTCCTGAAACCGTGTACATCAATTTGTGGGAAATCATAAAAGCAGCAACACCAACAAT 5520
        I  P  E  T  V  Y  I  N  L  W  E  I  I  K  A  A  T  P  T  I

5521  TATGAATAATCGATTGTCATCAGTTCATGCTGCCAAGGTTGCTTATACATTGAAGACAGA 5580
        M  N  N  R  L  S  S  V  H  A  A  K  V  A  Y  T  L  K  T  D

5581  TGTTCATTCAATCCCAAGAACTCTCGAGATCATTGATAGGCTTATTCAAGAAGAGCAAAC 5640
        V  H  S  I  P  R  T  L  E  I  I  D  R  L  I  Q  E  E  Q  T

5641  AAAACACGCGCATTTCCAAATGGTGAATTCTAATCTGGCGGGTCCGCACAATTTCTCTCT 5700
        K  H  A  H  F  Q  M  V  N  S  N  L  A  G  P  H  N  F  S  L
```

```
5701  CCTTGGTATAGCAAACTCAATAAGGCGCCGATACTTAGTCGATCATTCTATGCATAATCT 5760
        L  G  I  A  N  S  I  R  R  R  Y  L  V  D  H  S  M  H  N  L

5761  TGAACAGCTTCATCGTTGTCGAGCTCAAATCGCACAATATGCACAACTGGCGAGTGATCC 5820
        E  Q  L  H  R  C  R  A  Q  I  A  Q  Y  A  Q  L  A  S  D  P

5821  AAATGCGATTGACTTTATATCTCAATTCACCGATTTAAAACTTGTTCATTTTCAATCCAA 5880
        N  A  I  D  F  I  S  Q  F  T  D  L  K  L  V  H  F  Q  S  K

5881  ATACGAAATCGCAAAAGAGCTGAATCTCAGAGGTAAATGGAATGGTGGTAACATTACGAA 5940
        Y  E  I  A  K  E  L  N  L  R  G  K  W  N  G  G  N  I  T  N

5941  TGATCTACTGGTTGCCGGATTCGCGTTCATTGGTGGCATTTGGATGATATATAGCAAGTT 6000
        D  L  L  V  A  G  F  A  F  I  G  G  I  W  M  I  Y  S  K  F

6001  TAAGCAAATAACGCAAGAGCCAGTGCATTTCCAAGGAATGAATAAACGCCAACGACAAAA 6060
        K  Q  I  T  Q  E  P  V  H  F  Q  /G  M  N  K  R  Q  R  Q  K

6061  GCTACGTTTCCGGGACACACGAGATCGTAGACTCGGAAGGGAAATCTTTGCTGATGATGA 6120
        L  R  F  R  D  T  R  D  R  R  L  G  R  E  I  F  A  D  D  E

6121  AACAATGGGTGATTATTTTGGTTCAGCATATAAAGCCAAGAATAAGAGCAAAGGAACAAC 6180
        T  M  G  D  Y  F  G  S  A  Y  K  A  K  N  K  S  K  G  T  T

6181  AAAGGGAATGGGACACAAAACGCGTAAGTTTATACACATGTATGGAGTGGATCCCTCCGA 6240
        K  G  M  G  H  K  T  R  K  F  I  H  M  Y  G  V  D  P  S  E

6241  ATATGACACAGTGCGCTTTGTCGATCCATTAACGGGTGGTACACTCGATGAACAACCAAA 6300
        Y  D  T  V  R  F  V  D  P  L  T  G  G  T  L  D  E  Q  P  N

6301  CTGCGATATCTCTCTTGTTCAACAGCACTTTTCAGATATAAGACAACAGATGATCGACAA 6360
        C  D  I  S  L  V  Q  Q  H  F  S  D  I  R  Q  Q  M  I  D  N

6361  CGATGAATTGGATACGCAGCGTATTTTCTCGAATCCAGGAATAACGGCTTACTTCATTAG 6420
        D  E  L  D  T  Q  R  I  F  S  N  P  G  I  T  A  Y  F  I  R

6421  AAACGCTGCAAAAACCGCATTGAAAGTTGACTTGACACCACATAACCCTCGTCTTGTTTG 6480
        N  A  A  K  T  A  L  K  V  D  L  T  P  H  N  P  R  L  V  C

6481  TGCTAGAGCAGCTACCATATCAGGATTTCCAGAGCGCGAACTTGAGTTAAGGCAAACTGG 6540
        A  R  A  A  T  I  S  G  F  P  E  R  E  L  E  L  R  Q  T  G

6541  TATACCCGTTGAAATATCAAAGGACATCGTCCCAACAGCTAAGGAAACTGAGGTAGTGAT 6600
        I  P  V  E  I  S  K  D  I  V  P  T  A  K  E  T  E  V  V  M

6601  GTTTGAAAGCAAAGGACAAGTGAGTGGTCTTCAGGATTACAACGCTATATCAAGTGCTGT 6660
        F  E  S  K  G  Q  V  S  G  L  Q  D  Y  N  A  I  S  S  A  V

6661  CTGTCATCTGAAGAACACGAGTGAAGGTCATATCATTTCGACATTTGGAATTGGGTATGG 6720
        C  H  L  K  N  T  S  E  G  H  I  I  S  T  F  G  I  G  Y  G

6721  TAATACAATCATTGTGAATAAGCATCTCATGACTAACCAAAATGGAACACTATACATTAA 6780
        N  T  I  I  V  N  K  H  L  M  T  N  Q  N  G  T  L  Y  I  K

6781  AACACGATTTGGTGAATTCCAGATAGCTAGTATTCCAACTTTAAAGATGCTTCCCATACC 6840
        T  R  F  G  E  F  Q  I  A  S  I  P  T  L  K  M  L  P  I  P
```

```
6841  TGAACGTGATCTTCTCATGATTCAAATACCCAAAGACATGCCACCTTTCCCATCCAAATT 6900
        E  R  D  L  L  M  I  Q  I  P  K  D  M  P  P  F  P  S  K  L

6901  GAAATTTAGGCAACCTATGGAAGGTGAACGAATCTGCCTGGTTGGCTCACTATTCCAGAC 6960
        K  F  R  Q  P  M  E  G  E  R  I  C  L  V  G  S  L  F  Q  T

6961  GAACCATGTACGACAATTGGTTTCAGACTCATCACGATGCACAAGAATACCTGACACTCA 7020
        N  H  V  R  Q  L  V  S  D  S  S  R  C  T  R  I  P  D  T  H

7021  TTTTTTTAAGTACTGGATTTCGACAAAAATTGGACAATGTGGAACACCCATCGTCGGGAT 7080
        F  F  K  Y  W  I  S  T  K  I  G  Q  C  G  T  P  I  V  G  I

7081  AAAAGATGGGGCAATCGTTGGAATTCATAGCTTAGGTGCTAATTCTGGTATGGGAAATTA 7140
        K  D  G  A  I  V  G  I  H  S  L  G  A  N  S  G  M  G  N  Y

7141  CTTCACTGCATACCCGGAAAATTTCCAGCAAGACATATTGGAACGATTAGCGCAACTCGA 7200
        F  T  A  Y  P  E  N  F  Q  Q  D  I  L  E  R  L  A  Q  L  D

7201  TTGGACGAAGAAGTGGAGGTTCAACATAGATGGAATTTTGTGGGGAGCTTTGGGGTTGAC 7260
        W  T  K  K  W  R  F  N  I  D  G  I  L  W  G  A  L  G  L  T

7261  GCGTCTATCTAGTGATGGAATTTTAAACCTGTTAAGAAGATAGATCAGCTAACACTTGA 7320
        R  L  S  S  D  G  I  F  K  P  V  K  K  I  D  Q  L  T  L  D

7321  CATAGTCCATGAACAATCACATATTCTGGAACATTGGGTTGCAAATGCATTGCATACGAA 7380
        I  V  H  E  Q  /S  H  I  L  E  H  W  V  A  N  A  L  H  T  N

7381  TTTAAAAGCCGTTGCCCAGACGCCGAGTCAACTTATCACTAAACATGTTGTTAAAGGTAA 7440
        L  K  A  V  A  Q  T  P  S  Q  L  I  T  K  H  V  V  K  G  K

7441  ATGTGAGTATTTTGAACAATACTGCAGAATAAACGCAGATGCGAATGCCTTTTTCAAACC 7500
        C  E  Y  F  E  Q  Y  C  R  I  N  A  D  A  N  A  F  F  K  P

7501  ACTAGTTGGTGCATACGCTCCCACAGTCTTAAGCAAGGAAGCATACCTGAAAGACATCCT 7560
        L  V  G  A  Y  A  P  T  V  L  S  K  E  A  Y  L  K  D  I  L

7561  CAAATATGCGAGTGTAATCCCAGTCGGAAAAGTTAACACTGACTTGTTTGATGCAGCTGT 7620
        K  Y  A  S  V  I  P  V  G  K  V  N  T  D  L  F  D  A  A  V

7621  TGATGATCTAATTGCACACATGGAAAGAATTGGTTTCACCCAATGTGAATACATAACTGA 7680
        D  D  L  I  A  H  M  E  R  I  G  F  T  Q  C  E  Y  I  T  D

7681  TGCCGGAACAATCATTAACAGCTTGAATATGAACGCAGCAGTGGGATCATTGTTTGTTGG 7740
        A  G  T  I  I  N  S  L  N  M  N  A  A  V  G  S  L  F  V  G

7741  CAAGAAACGAGCGTACTTCGAAACATGCACAGCGCAAGATAGAGAAGATCTAGTTCAAGC 7800
        K  K  R  A  Y  F  E  T  C  T  A  Q  D  R  E  D  L  V  Q  A

7801  CAGCTGTAAGCGTCTATATTATGGTGAAAAAGGAGTCTGGAATGGATCCTTGAAAGCTGA 7860
        S  C  K  R  L  Y  Y  G  E  K  G  V  W  N  G  S  L  K  A  E

7861  ATTGAGACCTATTGAGAAAGTTGAATTAAAGAAAACTCGTACTTTCACAGCGGCACCCAT 7920
        L  R  P  I  E  K  V  E  L  K  K  T  R  T  F  T  A  A  P  I

7921  CGATACTTTACTTGGTGGAAAAGTGTGTGTTGATGATTTCAACCAGAAATTTTATAACCA 7980
        D  T  L  L  G  G  K  V  C  V  D  D  F  N  Q  K  F  Y  N  H
```

```
7981  CCACATTAAAGGTCCGTGGACCGTCGGGATGACGAAGTTCTACAAAGGGTGGGACGAACT 8040
        H   I   K   G   P   W   T   V   G   M   T   K   F   Y   K   G   W   D   E   L

8041  ATTAAACTGTCTGCCGGACAACTGGATCTATTGCGATGCCGATGGGAGCCAATTCGACAG 8100
        L   N   C   L   P   D   N   W   I   Y   C   D   A   D   G   S   Q   F   D   S

8101  TTCGATCACACCATACATCATGAATGCAGTTTTGAAGGTACGTGAGCATTTTATGGAAGA 8160
        S   I   T   P   Y   I   M   N   A   V   L   K   V   R   E   H   F   M   E   D

8161  CTGGGACCTTGGGAAGGAGATGTTACGAAATCTTTACACTGAGATAGTGTATACGCCTAT 8220
        W   D   L   G   K   E   M   L   R   N   L   Y   T   E   I   V   Y   T   P   I

8221  ATCTACACCAGATGGCACGATAATCAAGAAACACAAAGGGAACAATAGTGGCCAACCATC 8280
        S   T   P   D   G   T   I   I   K   K   H   K   G   N   N   S   G   Q   P   S

8281  TACAGTCGTAGATAATTCACTAATTGTGTGCATCGCTGTGAAATATGCGCTCTTAAAAGA 8340
        T   V   V   D   N   S   L   I   V   C   I   A   V   K   Y   A   L   L   K   E

8341  AGGATTGAAACCAGAAGAGTTTGATAACGTGATCCGTTTCTTTGTTAATGGAGACGATCT 8400
        G   L   K   P   E   E   F   D   N   V   I   R   F   F   V   N   G   D   D   L

8401  TCTCATAGCCGTGAAACCATCCCATGAGCACATCTTGAACAACTTCTCAGAATACTTCGC 8460
        L   I   A   V   K   P   S   H   E   H   I   L   N   N   F   S   E   Y   F   A

8461 TGAGCTTGGATTGAACTACGAATTCGACAACCGACACACTGACAAATCCAGACTTTGGTT 8520
       E   L   G   L   N   Y   E   F   D   N   R   H   T   D   K   S   R   L   W   F

8521  TATGTCCCACTGTGGCATTCGTGATGAAGATATCTGGATCCCGAAACTTGAACCCGAACG 8580
        M   S   H   C   G   I   R   D   E   D   I   W   I   P   K   L   E   P   E   R

8581  CATAGTGTCTATCCTTGAATGGGATAGAAGTAAATTGCCTGAGAATCGTCTGGAAGCAGT 8640
        I   V   S   I   L   E   W   D   R   S   K   L   P   E   N   R   L   E   A   V

8641  GTGTGCCGCCATGGTTGAGGCGTGGGGCTATAATGATTTACTTCACCAAATACGGTTGTT 8700
        C   A   A   M   V   E   A   W   G   Y   N   D   L   L   H   Q   I   R   L   F

8701  TTACCAATGGCTCCTCGAACAAGAGCCATTCAACGATCTGGCCGCAAATGGGTATGCACC 8760
        Y   Q   W   L   L   E   Q   E   P   F   N   D   L   A   A   N   G   Y   A   P

8761  GTACATTAGTGAAGTAGCATTGCGCGCGCTGTATACAGGAAGCCAACCAACCAATACATC 8820
        Y   I   S   E   V   A   L   R   A   L   Y   T   G   S   Q   P   T   N   T   S

8821  AATGCTACCATATTTACATCAACTTGTCGATGATTTTGAACATGATGCAGACGAAGATTA 8880
        M   L   P   Y   L   H   Q   L   V   D   D   F   E   H   D   A   D   E   D   Y

8881  TGAAGTTCATTTCCAGAGTGGTGAAAAAGAAAAGGAGAAAAATTGATACGGAGGAGGCTAA 8940
        E   V   H   F   Q   /S  G   E   K   E   K   E   K   I   D   T   E   E   A   K

8941  GCGTACAATAAAGAAATTGAAAGAGGTTGTCGATGCAGGCAATGAGCAATCTATTACTAC 9000
        R   T   I   K   K   L   K   E   V   V   D   A   G   N   E   Q   S   I   T   T

9001  ATCCTCGAAAGTTAAGCAGGACCAAATAAGTACCGAAAAGGACAAAGATGTTTCAACCTC 9060
        S   S   K   V   K   Q   D   Q   I   S   T   E   K   D   K   D   V   S   T   S

9061  GACGCGTGGCACATTTGCAATTCCAAAACTGAGATTGATGAATCAAGTCATGAAACTACC 9120
        T   R   G   T   F   A   I   P   K   L   R   L   M   N   Q   V   M   K   L   P
```

```
9121  AAAGATTGGAAATAGAAATTTATTAAATCTCGAACACTTACTCACGTATAAGCCTGAACA 9180
        K  I  G  N  R  N  L  L  N  L  E  H  L  L  T  Y  K  P  E  Q

9181  AGTTGACATATCCAATACGCGCGCAACAAAGCAACAGTTTGAAACTTGGTGTGAACGTGT 9240
        V  D  I  S  N  T  R  A  T  K  Q  Q  F  E  T  W  C  E  R  V

9241  TCGTGATGCGTATGGATTGAAAGATGATGAAATGAGTATAGTGATGAATGGATTTATGGT 9300
        R  D  A  Y  G  L  K  D  D  E  M  S  I  V  M  N  G  F  M  V

9301  CTGGTGCATTGAAAATGGCACATCACCCAATGTCAATGGGGTCTGGACTATGATGGATGG 9360
        W  C  I  E  N  G  T  S  P  N  V  N  G  V  W  T  M  M  D  G

9361  CACAGAACAAAACGAATATCCTTTGAAACCCATGGTTGAAAATGCGAAACCTACCCTCCG 9420
        T  E  Q  N  E  Y  P  L  K  P  M  V  E  N  A  K  P  T  L  R

9421  CCAAATCATGGCGCATTTCTCTGACGCTGCGGAAGCTTACATTGAAATGCGGAACACAAA 9480
        Q  I  M  A  H  F  S  D  A  A  E  A  Y  I  E  M  R  N  T  K

9481  GGAACCATATATGCCAAGATATGGACAGCAGCGTAACCTAACCGACATGAGCCTTGCACG 9540
        E  P  Y  M  P  R  Y  G  Q  Q  R  N  L  T  D  M  S  L  A  R

9541  ATGCGCTTTCGATTTTTATGAAATGACATCTCGAACACCAGCACGTGTCCGAGAAGCTCA 9600
        C  A  F  D  F  Y  E  M  T  S  R  T  P  A  R  V  R  E  A  H

9601  TATTCAAATGAAAGCAGCTGCTCTGCGTAATACAAGAACTAGAATGTTTGGACTTGATGG 9660
        I  Q  M  K  A  A  A  L  R  N  T  R  T  R  M  F  G  L  D  G

9661  CAATGTTGGCACTCAAGAGGAGGATACTGAGAGACACACCACTGAGGACGTTAATAGGAA 9720
        N  V  G  T  Q  E  E  D  T  E  R  H  T  T  E  D  V  N  R  N

9781  CATGCATAATCTTCTTGGTCTCAAAATGTAATACTATATAATGCATGTGTGGTGAACTAT 9780
        M  H  N  L  L  G  L  K  M

9781  CCTATATATAGCAGTATCTGTATGTTTATTCATCTCCTCTCCCTTGCAGACCCTGATATC 9840

9841  CTGTAGTGTGGGTGGCCCACCAGCGCTTAGGCAGCCAGGACAGGTTCTATTCTGTAAGTG 9900

9901  CTCTTCTGCTATGTTGCCTAGGTTAGGAGTAAAAGTCTATCGCATGCGTGTGGTAGGCTT 9960

9961  CCCGCGATCCTATGTTTCATAGTAGCTTTGAGTGAAAAAAAAAAAAAAAAAAAAAA   10017
```

		Potybox				
		UCAACACAACAU				
		Box a		Box b		
VIRUS	nt	ACACAACAU	nt	UCAAAGCA	nt	IC
WLMV	6	UCAACACAACAU	48	ACAAGCAA	19	AUG
PRSV	12	UCAACACAACAU	52	ACAAGCAA	2	AUG
JGMV	1	CCAACACAACAC	108	AUAAGCAG	6	AUG
SbMV	21	ACAACA--AACAA	37	ACAAGCAA	53	AUG

Fig. 3-3. Alignment of the 5'untranslated region (5'UTR) of watermelon mottle virus (WLMV) and of three other potyviruses. Variations among distinct viruses in the sequences of the potybox, and box a and box b, and in the number of nucleotides surrounding these consensus regions is shown. Nucleotides that are distinct from the consensus are underlined.
IC = initiation codon; nt = nucleotides.
Papaya ringspot virus (PRSV); potato virus Y (PVY); Johnsongrass mosaic virus (JGMV); soybean mosaic virus (SbMV). (Figure was adapted from Fig. 4-1 in Shukla *et al.*, 1994).

	P1/HC-Pro	KITC	FRNK	PCCC	PTK
WLMV	AQY/S.CG..PSH.	RITC.	DKC.	FRNKATS.	CIVTKHNQYLHPCCC. VT. PTKNHLVI
PRSV-HA	EQY/N.CG..PCH	KITC	TKC.	FRNKRSG.	CIALRNNNYVHVCSC. VT. PTKNHIVL
ZYMV-CA	DHS/S.CG..PVV	KLSC	QTC.	FRNKRSS.	CVSQQNGNYIHSCCC. VT. PTKRHLVI
ZYMV-RU	DHY/S.CG..PVV	KLSC	QTC.	FRNKRSS.	CVSQQNGNYIHSCCC. VT. PTKRHLVI
PPV-RA	IHY/S.CG..PCG	KITC	LQC.	FRNKISS.	CTSKRGENFVYQCCC. VT. PTKNHLVV
TuMV	VHF/S.CG..PCG	KITC	PDC.	FRNKISQ.	CVSKLQGDFVHACCC. VT. PTKHHLVI
TVMV	THF/S.CG..PCG	KITC	KRC.	FRNKISS.	CVSKLQGDFVHACCC. VT. PTKNHLVI
TEV	THY/S.CG..PCG	KITC	KRC.	FRNKISS.	CISMRNGNYVYPCCC. VT. PTKRHLVI
PSbV	DHF/S.CG..PCW	RITC	TKC.	FRNKISA.	CVSRRNGDFVYPCSC. VT. PTRNHLVI
PStV	HHY/S.CG..PVK	KLSC	TKC.	FRNKRSS.	CVAKQNGNFVYPCCC. VT. PTKRHLVV

Fig. 3-4. Amino acid sequence alignment of the helper-component protease protein (HC-Pro) of watermelon leaf mottle virus (WLMV) compared to selected potyviruses. The cleavage site between the P1 protein and the helper-component protease (P1/HC-Pro) and highly conserved motifs within the HC-Pro of all viruses in the *Potyvirus* genus (KITC, FRNK, PCCC and PTK) are indicated above the alignment.
Hawaii strain of papaya ringspot virus (PRSV-HA), zucchini yellow mosaic virus strain California (ZYMV-CA), Reunion Island strain of zucchini yellow mosaic virus (ZYMV-RU), Rankovic strain of plum pox virus, (PPV-RA) turnip mosaic virus (TuMV), tobacco vein mottle virus (TVMV), tobacco etch virus (TEV), pea seed borne mosaic virus (PSbMV), peanut stripe virus (PStV). Sequences were obtained from GenBank, alignments were made using CLUSTAL W (1.7).

CS	VXXQ/G,S,A
PNLS	KRQRQKLRFRDTRD
WLMV	VHFQ/GMNKRQRQKLRFRDTRDRRLGREIF-FG-SAYKAKNKSKG--TTKGMGHKT-RKFIHMYGVDPSEYDT
PRSV-HA	VFHQ/GFSARQRQKLRFKSAANAKLGREVY-FG-EAYTKKGNKKG--KMHGMGVKT-RKFVATYGFKPEDYSY
ZYMV-CA	VRVE/SKKR-RSQKLKFRDAYDRKVGREIF-FG-EAYTKRGKVKGNNNTKGMGRKT-RNFVHLYGVEPENYSF
ZYMV-RU	VRVE/SKKR-RSQKLKFRDAYDRKVGREIF-FG-EAYTKRGKVKGNNSTKGMGRKT-RNFVHLYGVEPENYSF
PStV	VTTQ/GKKR-MIQKLKFRDAFDRKVGREVY-FG-EAYTKKGKQKGSTKTKGMGRKT-RNFVHMYGVEPENYSM
PPV-RA	VVHQ/GFNRRQRQKLKFRQARDNRMAREVY-FG-SAYSKKGKSKG--KTRGMGTKT-RKFVNMYGYDPTDYNF
TuMV	VTHE/AKGKRQRQKLKFRSIRDNKMGREVY-FG-DAYTKKGKSKG--RTRGLGHKN-RKFINMYGFDPEDFSA
TVMV	VKFQ/GK---SRRRLQFRKARDDKMGYIMH-FG-AAYTKKGKSKG--KTHGAGTKA-HKFVNMYGVSPDEYSY
TEV	VYFQ/GKKN-QKHKLKMREARGARGQYEVA-FG-SAYNNKGKRKG--TTRGMGAKS-RKFINMYGFDPTDFSY
PSbMV	VKLE/GKS--KAKTLRFRQARDNNAKYEVF-FG-EAYTKKGKKSG--KARGMGVKT-KKFVNVYGFDPCEYSL
CONSENSUS	V--Q/-K-----q-Lkfr-ard----------AYtKKGk-kG-----G---K----FvnmYG--P--ys-

Fig. 3-5. Amino acid sequence alignment of watermelon leaf mottle virus (WLMV) and selected potyviruses. The consensus sequence for the cleavage site (CS) between the 6K2 protein and the VPg-Pro of potyviruses and the putative nuclear localization signal of WLMV (PNLS) are indicated above the alignment. The consensus around the tyrosine residue that links the VPg to the viral RNA is underlined. The potyviral amino acid consensus sequence is indicated below the sequence alignment (CONSENSUS).
Hawaii strain of papaya ringspot virus (PRSV-HA); zucchini yellow mosaic virus strain California (ZYMV-CA); Reunion Island strain of zucchini yellow mosaic virus (ZYMV-RU); Rankovic strain of plum pox virus (PPV-RA); turnip mosaic virus (TuMV); tobacco vein mottle virus (TVMV); tobacco etch virus (TEV); pea seed borne mosaic virus (PSbMV); peanut stripe virus (PStV). Sequences were obtained from GenBank, and alignments were made using CLUSTAL W (1.7).

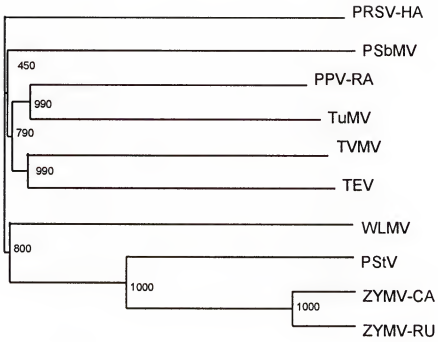

Fig. 3-6. Neighbor-joining tree calculated by CLUSTAL W (1.7) showing the relationship of WLMV to selected potyviruses using full-length amino acid sequences. Numbers represent bootstrap values out of 1,000 replications.
Papaya ringspot virus Hawaii strain (PRSV-HA); pea seed-borne mosaic virus (PsBMV); plum pox virus Rankovic strain (PPV-RA); turnip mosaic virus (TuMV); tobacco vein mottling virus (TVMV); tobacco etch virus (TEV); watermelon leaf motle virus (WLMV); peanut stripe mosaic virus (PStV); zucchini yellow mosaic virus CA strain (ZYMV-CA); zucchini yellow mosaic virus Reunion Island (ZYMV-RU). Sequences were selected from the output of BLASTP and were obtained from GenBank.

CHAPTER 4
SEROLOGICAL CHARACTERIZATION OF WATERMELON LEAF MOTTLE VIRUS (WLMV)

Introduction

Watermelon leaf mottle virus (WLMV) was isolated from watermelon leaf samples in Florida in 1990, but since then, no additional isolates have been found in limited testing of squash and watermelon field samples from Florida (Purcifull *et al.*, 1998). WLMV is transmitted nonpersistently by *Myzus persicae* (Sulzer) to squash and watermelon. However, the efficiency of transmission is yet to be determined. No other aphid species have been tested as vectors of WLMV. The distribution and epidemiology of the virus and the source of the initial infection are unknown. WLMV is of potential importance because it causes mosaic and mottle in various cultivars of squash and watermelon, and mostly symptomless infections of cucumbers and melons. It also causes mosaic in 'Prelude II' (Purcifull *et al.*, 1998), a cultivar of squash carrying coat protein mediated resistance to zucchini yellow mosaic virus (ZYMV) and watermelon mosaic virus 2 (WMV-2)

Several potyviruses have been described infecting cucurbit crops in many parts of the world, but most have only been partially characterized and many have not been sequenced to date (Table 4.1). Of the 18 viruses listed in Table 4-1, many remain as tentative species of the *Potyvirus* genus, and seven are not listed by the ICTV in its Sixth

Report (Murphy *et al.*, 1995). Some of the viruses named may be strains of the same virus, rather than new species. According to Shukla *et al.* (1994), because of the significant difference in the sequence identities found between distinct viruses and strains, the status of a potyvirus as a species or strain can be readily established by comparing sequence data of the coat protein gene. However, alternative methods need to be used when sequence information is not available, and it may be more practical to use classical virological methods based on the biological properties for initial comparisons of partially characterized or newly isolated viruses, particularly when sequencing facilities are not immediately available. Host range and symptomatology are frequently used for the initial identification of a newly isolated virus, although they may be unreliable as definitive identifiers of most potyviruses. An established set of differential hosts is useful for differentiating distinct and well characterized potyviruses with overlapping host ranges, and host range and symptomatology have been considered as the most useful approach for the differentiation of potyvirus strains and pathotypes (Shukla *et al.*, 1994).

The most efficient and widely used procedures for virus identification are those based on the antigenic properties of the virus coat protein. Among serological methods, SDS immunodifusion is one of the most useful diagnostic techniques for the identification and classification of potyviruses; it is considered to be one of the most specific and the least sensitive of the serological techniques (Shukla *et al.*, 1994). There are several highly sensitive serological techniques and among them double antibody sandwich enzyme-linked immunosorbent assay (DAS ELISA), radioimmunoassay (RIA) and immune-specific electron microscopy (ISEM) are the most specific, whereas, indirect ELISA and Western blots have broad specificities (Shukla *et al.*, 1994). Indirect ELISA

is frequently used in field surveys of plant viruses, as it can be used to test large numbers of samples and does not require that primary antibodies be conjugated to the enzyme. It is considered the best method for detecting serologically distinct strains of a virus (Van Regenmortel, 1982). Although DAS ELISA is more specific (Shukla *et al.*, 1994), its strain specificity may be a drawback for diagnostic work (Van Regenmortel, 1982).

The objectives of the research described in this chapter were:

1 - To identify plants that gave clear and consistent symptoms when inoculated with WLMV and thus may be used as indicator hosts for diagnosis and for differentiation of WLMV from other potyviruses infecting cucurbits.

2 - To express the coat protein of WLMV in *Escherichia coli*, and to produce polyclonal antisera in rabbit and in mice to the expressed coat protein.

3 - Characterize the polyclonal antisera and test their use in diagnosis and identification of WLMV, using SDS immunodiffusion, indirect ELISA and Western blotting.

4 - Test the commercially available monoclonal antiserum PTY-1 and an admixture of monoclonal antisera (ADMIX) for detection of WLMV.

Materials and Methods

Virus Isolates

All the virus isolates used were maintained in squash (*Cucurbita pepo*) 'Early Prolific Straight Neck Squash' (EPSN). Two subcultures of the original isolate of WLMV were used, a culture maintained exclusively by mechanical transmission, and an aphid-transmitted culture. The aphid-transmitted culture was transmitted by aphids to

EPSN, and preserved by desiccation in dried infected plant tissues over Silica Gel at 4°C. When needed, it was recovered by mechanical transmission to EPSN followed by aphid transmission to EPSN. Leaves showing symptoms were collected and used in the experiments. Other viruses used for comparisons with WLMV were: papaya ringspot virus type W (PRSV-W), isolate W-1A (Purcifull and Hiebert, 1979); zucchini yellow mosaic virus (ZYMV) (isolate FC 1119 AT of Purcifull *et al.*, 1984; Wisler, 1992); watermelon mosaic virus 2 (WMV-2) (isolate FC 1656 of Purcifull), squash mosaic virus (SqMV) (isolate FC 979 of Purcifull). Moroccan watermelon mosaic virus (MWMV) (MOR-ATCC PV-957, a subculture of the isolate used by Purcifull and Hiebert [1979] which was originally from Lockhart [Fischer and Lockhart, 1974]). MWMV was maintained under quarantine conditions at the Department of Plant Industry (DPI) (Table 4-2).

Aphid Transmission of WLMV

A colony of *Myzus persicae* was maintained on Indian mustard (*Brassica juncea* L.) 'Florida Broad Leaf' and aphids were used for transmission of the virus. Four pots containing three seedlings of EPSN squash each, were used for each experiment. Aphids were starved for 1 h and were allowed acquisition access on an infected squash leaf for 15 min. Five aphids per plant were transferred to the cotyledons of each plant, and allowed to feed overnight. The leaf used for acquisition-access was used for mechanical inoculation of cotyledons of three squash seedlings per pot in four pots.

The aphids were killed by spraying with Enstar II Kinopren, or 'Safe Soap' on the following day, and the plants were transferred to a growth room or to a greenhouse, and observed for symptom development. Symptomatic leaves were dried down and

maintained at 4°C under vacuum and over desiccant for long-term maintenance of the aphid-transmitted isolate.

Inoculation of Test Plants

Mechanical inoculation of test plants was performed by rubbing a swab of cheesecloth dipped in leaf extract prepared by grinding symptomatic leaves of EPSN infected with WLMV, with a pestle and mortar in potassium phosphate buffer (20 mM, pH 7.5) and 600 mesh Carborundum. Leaves were rinsed after inoculation, and plants were maintained for two to six weeks in a greenhouse at 20°C to 24°C for observation of symptoms. Inoculations were performed on a minimum of five plants, and the experiment was conducted at least two times. Leaf samples of all plants were collected, pooled and mechanically inoculated on the cotyledons of six to twelve EPSN plantlets as described above for virus recovery assays. Some test plants were also assayed by indirect ELISA and Western blots.

For host range tests, plants obtained from the USDA, ARS, North Central Regional Plant Introduction Station (NCRPIS) were: *Cucumis metuliferus* E. Mey. ex Naud. PI 292190, PI 482459; *Cucurbita okeechobeensis* var. *okeechobeensis* (Small) L.H.Bailey; *Cucurbita moschata* (Duch. ex Lam.) Duch. ex Poir; *Lagenaria siceraria* (Molina) Stand; *Luffa aegyptiaca* P. Mill; *Momordica charantia* L. *Cucumis metuliferus* PI 292190 and Acc. 2459 were obtained from R. Provvidenti.

Other cucurbits tested were: *Citrullus lanatus* (Thunb.) Matsumura & Nakai 'Crimson Sweet', 'Charleston Gray'; *Cucurbita pepo* L. 'Early Prolific Straightneck Squash' (EPSN), and 'Prelude II'; *Cucumis sativus* L. 'Surinam', 'TGM-1 Taiwan' (from R. Provvidenti); *Cucumis melo* L. 'Hale's Best Jumbo'; *Cucurbita ecuadorensis* Cutler &

Whitaker; and *Luffa acutangula* (L.) Roxb. Noncucurbitaceous plants tested were *Cassia occidentalis* L.; *Chenopodium amaranticolor* Coste and Reyn.; *C. quinoa* Willd.; *Phaseolus vulgaris* L. 'BT-1', 'BT-2' and 'Top Crop'; *Nicotiana benthamiana* Domin.; *Nicotiana tabacum* L. 'Burley 21' and 'NN'.

Amplification and Expression of the Coat Protein Coding Region

The coat protein (CP) coding region was amplified from a sequenced clone pCPB (Chapter 2) by the polymerase chain reaction (PCR). A forward primer 5' TCTAGCTAGC**ATG**AGTGGTGAAAAAGAAAAGG 3', with an Nhe I restriction site and an 'ATG' initiation codon, and a reverse primer 5' ATCTAAAC**TTA**CATTTTGAGACCAAGAAG 3', complementary to the end of the CP coding region were used. PCR was performed in a UNO-Thermocycler (Biometra Inc., Tampa, FL). The following cycling program was used: 94°C for 2 min, 60°C (-1 C) for 1 min, and 72°C for 1 min, for ten cycles; followed by 25 cycles of 92° C for 1 min, 50° C for 30 sec and 72°C for 1 min; and an extension step of 5 min at 72°C.

The PCR product was purified by electrophoresis in 1% agarose. A band of approximately 0.9 kb was excised and the DNA was extracted from the agarose using a GEN-Elute agarose spin column (Supelco, Bellefonte, PA) equilibrated with 100 µl TE buffer (10 mM Tris-HCl, 1 mM EDTA, pH 8.0). The column was centrifuged at 16,000 *g* for 10 min at 4°C, and the column eluate was mixed with 2.5 volumes of 100% ethanol and 0.5 volume of 7.5 M ammonium acetate and incubated overnight at –20°C. The DNA was pelleted by centrifugation for 30 min at 16,000 *g*, and the pellet was washed with 70% ethanol. The pellet was vacuum dried and ressuspended in MBG water. The purified DNA was ligated into pGEM T vector (Promega, Madison, WI) by the 'A' overhangs to

generate pGEM-T-CP, according to manufacturer's instructions. *E. coli* DH5α was transformed with the ligation reactions essentially according to Studier *et al.* (1990). Transformed DH5α cells were plated onto LB/AMP plates (100 μg/ml) with IPTG (1 mM) and X-gal (0.08 μg/ml) for blue/white selection, and the plates were incubated overnight at 37°C. Positive single colonies were grown in LB/AMP (100 μg/ml) broth. DNA was purified by alkaline lysis miniprep, and screened for correct insert size by electrophoresis and digestion with EcoR I.

The plasmid pGEM-T-CP was double digested with Nhe I and Sac I (Life Technologies, Gaithersburg, MD) and the coat protein DNA fragment was ligated into the Nhe I and Sac I restriction sites of the polylinker of pETH 3c (McCarty *et al.*, 1991; Studier *et al.*, 1990) to generate the plasmid pETH-CP. *E. coli* strain BL21 (DE3) pLysS (Stratagene, La Jolla, CA) was used as the expression host and was transformed with the resulting ligation products. To confirm sequence and reading frame of the PCR amplified coat protein gene, pETH-CP DNA was purified from the transformed BL21 (DE3) pLysS and sequenced.

For expression of the recombinant protein, *E. coli* BL21 (DE3) pLysS was grown in M9 liquid culture medium (Sambrook *et al.*, 1989) with ampicillin (100 μg/ml) and chloramphenicol (25 μg/ml), and was induced with IPTG to a final concentration of 0.4 mM. Cells were harvested 3 h later by centrifugation at 5,000 *g*, ressuspended in TE buffer (10 mM Tris-HCl, pH 8.0, 1 mM EDTA), and frozen overnight at -80°C. The cell suspension was homogenized for 30 sec with an ISCO homogenizer, centrifuged at 10,000 *g* for 10 min, and the pellet was ressuspended in a small volume of TE buffer.

Bacterial lysates carrying induced and non-induced recombinant plasmid cultures were analyzed by electrophoresis in 10% sodium dodecyl sulphate polyacrylamide gels (SDS-PAGE), followed by Coomassie blue staining or Western blotting. Western blots were performed as previously described (Towbin *et al.*, 1979), using a BioRad Mini PROTEAN II electrophoresis cell and a BioRad electrophoretic transfer cell (Bio-Rad Laboratories, Hercules, CA). Proteins separated on 10% SDS-PAGE were transferred to nitrocellulose membranes (0.45 μ pore size, Bio-Rad) by electrophoretic transfer at a constant voltage of 100 V for 1 h. Membranes were washed three times for 10 min in TBST (20 mM Tris-HCl, pH 8.2, 15 mM NaCl, 0.1% Tween 20), and blocked for 30 min at RT with *E. coli* lysate (1 mg/ml) in Blotto solution (5% nonfat dry milk in TBST). Membranes were washed in TBST as above, and incubated in AS 1163 diluted 1:2,000 for 2 h at RT or overnight at 4°C. After washing with TBST, the membranes were incubated in goat anti-rabbit IgG-conjugated alkaline phosphatase (Sigma Chemical Co., St. Louis, MO) for 2 h at RT, and were treated with developing solution as described above.

Bacterial lysate samples for 10% SDS-PAGE were prepared by (1:1) addition of ES buffer (75 mM Tris-HCl, pH 6.1 with 9 M urea, 7.5% β - mercaptoethanol and 4.5% SDS) (Rodriguez-Cerezo and Shaw, 1991). Samples were vortexed, boiled for 2 min and then centrifuged at 3,000 *g* for 3 min. Induced and non-induced recombinant bacterial cultures were sampled. Bacterial lysates were also mixed with equal volumes of 3% SDS and water and tested by immunodiffusion (Purcifull, 1990) against AS 1163. In these tests purified virus was used as a positive control and non-induced BL21 (DE3) pLysS was used as a negative control.

The expressed coat protein was purified by preparatory 10% SDS PAGE. The gel was stained with a cold solution of 0.2 M KCl for 7 min at 4°C, washed three times in cold distilled water, and the band of interest was excised and frozen overnight. The coat protein was electro-eluted from the gel slice using a Bio-Rad electroeluter at 10 mA/tube, with constant current for 5 h. The protein was collected and dialyzed overnight at 4°C. Spectrophotometric readings of dialyzed proteins were used to calculate protein concentration and protein aliquots were lyophilized and stored at -20°C.

Production of Polyclonal Antiserum

The expressed coat protein was used to produce a polyclonal antiserum (AS 1216) in a New Zealand white rabbit by a series of three weekly intramuscular injections, followed later by a booster injection. Preimmune serum was collected one week before the rabbit was immunized. The injections consisted of 3.6 mg of purified coat protein in Freund's complete adjuvant, followed by two injections of 3.6 mg and 1 mg protein in incomplete Freund's adjuvant. Serum from this rabbit (AS 1216) was collected once a week for eight weeks and a booster injection (2 mg) was given to the rabbit 11 weeks after the first injection. Antisera collected before and after the booster injection were used in serological studies.

The AS 1216 was titered against purified expressed coat protein (100 ng/100 µl), EPSN squash infected with WLMV, and healthy squash, using indirect ELISA essentially according to the plate-bound antigen protocol (Converse and Martin, 1990).

Polyclonal Antisera and Monoclonal Antibodies

Two polyclonal antisera produced in rabbits against WLMV coat protein were used: AS 1216 prepared against the expressed coat protein, and AS 1163 prepared against

the dissociated viral coat protein obtained from sucrose gradients (Purcifull *et al.*, 1998). Polyclonal antisera against PRSV-W, ZYMV, WMV-2, MWMV, were obtained from Purcifull, and are listed in Table 4.2. Purified immunoglobulin (IgG) of AS 1163, and polyclonal ascitic fluid were used in some tests. A polyclonal antiserum prepared to the cylindrical inclusion of WLMV (Purcifull and Hiebert, unpublished) was also used.

Monoclonal antibodies tested were PTY-1 (Agdia, Inc., Elkhart, IN), and Admix (an admixture of PTY-1, PTY-2, PTY-4, PTY-8, PTY-10, PTY-21, PTY-24 and PTY-30) provided by R.L. Jordan (Jordan and Hammond, 1991).

Purification of Immunoglobulin

The purification of Immunoglobulin G (IgG) from AS 1163, was done essentially according to Clark and Adams (1977). Leaves of healthy EPSN squash were extracted (1:10) in phosphate-buffered saline (PBS) and the extract was dialyzed against PBS overnight at 4°C. AS 1163 was mixed with the dialyzed healthy plant extract (1:10) and stirred for 4 h at 37°C and overnight at 4°C. The antiserum and plant extract mix was centrifuged at 5,000 g for 10 min, frozen overnight, thawed, and passed through a DEAE-Sephacel column.

Polyclonal Ascitic Fluid

Polyclonal ascitic fluid against WLMV, was prepared at the Hybridoma Core Laboratory, ICBR, University of Florida. The expressed coat protein was diluted (10 µg/100 µl) in PBS (Gibco BRL, Gaithersburg, MD) and emulsified with 100 µl RIBI MPL + TDM Emulsion (0.5 mg Monophosphoryl Lipid A, 0.5 mg synthetic Trehalose Dicorynomycolate, 40 µl Squalene [hexamethyl-tetracosahexane], 4 µl Monooleate [Tween 80]) (RIBI ImmunoChem Research Inc., Hamilton, Montana). Two BALB/C

mice were injected subcutaneously with 50 µl at each ventral groin site and 100 µl on the scruff of the neck. A third mouse was injected with 100 µl PBS emulsified with 100 µl RIBI MPL + TDM, but without virus coat protein, to produce ascitic fluid free of WLMV antibodies for use as a negative control in ELISA. The mice were immunized in the same way two weeks after the first injection. A test bleeding was performed a week later to check for antibody class switching and titer and a third injection was applied to the mice. Mice were test bled for the second time after another week, and a booster immunization was given to the mice as described above.

A third test bleeding was performed and each mouse was primed with 0.5 ml Pristane (2,6,10,14 – Tetra-Methyl Pentadicane) (Sigma Chemical Company, St. Louis, MO). One week later the mice were injected intraperitoneally with nonfused Sp2/0 cells.

Test bleedings were titered by plate-bound antigen indirect ELISA according to the procedure adopted by the Hybridoma Lab., using as primary antibodies antiserum from immunized mice and from the nonimmunized mouse; AS 1216 and pre-immune (PI 1216) rabbit serum were used as controls. All incubations proceeded for 1 h at 37°C. Flat bottom 96 well ELISA plates were coated with 50 µl purified expressed coat protein (5 µg/ml) diluted in PBS (Gibco BRL, Gaithersburg, MD), incubated and washed four times with PBST (PBS containing 0.05% Tween 20, pH 7.4). Plates were blocked with blocking buffer (PBS, 1.0 % bovine serum albumin [BSA], 0.02% sodium azide), incubated and washed with PBST. Wells were then coated with the diluted primary antibody, incubated and washed. Secondary antibodies were rabbit anti-mouse IgG whole molecule alkaline phosphatase conjugate, goat anti-mouse IgG gamma chain specific alkaline phosphatase conjugate and goat anti-mouse IgM µ chain specific alkaline

phosphatase conjugate (Sigma Chemical Company, St. Louis, MO). Enzyme substrate solution (p-nitrophenyl phosphate in substrate buffer at a concentration of 1 mg/ml, pH 9.6) was added to the plates after the last wash. The plates were read at A_{405}, and absorbance values recorded at 30 min intervals. Ascitic fluid was collected from the mice, centrifuged to remove cells, and stored frozen.

Serological Tests

Plates for immunodiffusion (0.5 % Noble agar, 1% sodium azide and 0.5% SDS) were prepared according to Purcifull (1990). Antigens for immunodiffusion were prepared by macerating 1 g of leaf tissue in 1 ml of distilled water and 1 ml of a 3% SDS solution (Purcifull, 1990). The crude extract was expressed through cheesecloth, and used immediately.

Indirect ELISA with plate-bound antigen was performed as previously described (Converse and Martin, 1990) with modifications. Microtiter plates were loaded with 100 µl volume of each reagent in each step, and incubations were done in a moist chamber for 1 h at 37°C. Samples were extracted in extraction buffer (0.13% Na_2SO_3, 2% polyvinyl pyrrolidone [PVP - 40], 0.2% BSA in PBST, pH 9.6), and added to duplicate or triplicate wells and the plates were incubated for 1 h at 37°C. Plates were washed with PBST three times for 3 min. Polyclonal antisera or monoclonal antisera diluted in conjugate buffer (2% PVP-40, 0.2% BSA in PBST) were added to the wells and allowed to incubate. Plates were washed and loaded with whole molecule alkaline phosphatase-labeled goat anti-rabbit IgG or anti-mouse IgG diluted in conjugate buffer (1:30,000). Plates were incubated and washed as before. Enzyme substrate solution (para-nitrophenyl phosphate in substrate buffer at a concentration of 1 mg/ml) was added to the plates, and they were

incubated at room temperature in the dark for color development. Absorbance values (A_{405}) were determined using a microplate reader (Bio-Tek Instruments, Winooski, VT) at intervals varying from 15 min to 2 h. Plants giving absorbance values above two times those of noninoculated plants were considered hosts.

For triple antibody sandwich indirect ELISA (TAS indirect ELISA), plates were coated with 100 µl polyclonal ascitic fluid (1:1,000), incubated for 1 h at 37°C, and washed with TBST buffer three times. Leaf extracts were added to wells and subsequent procedures were conducted as for indirect ELISA.

Samples for Western blots were prepared by macerating leaf tissue in liquid nitrogen, Laemmli dissociating solution (20% 1.25 M Tris-HCl, pH 6.8, 2.5% SDS, 4% 2-mercaptoethanol, 10% sucrose) (Laemmli, 1970) was added at a concentration of 1:1 (w:v). Leaf extracts were expressed through cheesecloth and transferred to microfuge tubes that were placed in boiling water for 3 min. The tubes were centrifuged for 3 min at 12,000 *g*, and the supernatant was loaded onto the gels.

To check for the presence of WLMV coat protein in host plants, indirect ELISA was performed using extracts from some of the plants tested in the host range study, using AS 1216 collection number 13 (AS 1216 # 13) diluted 1:8,000, and purified IgG from AS 1163. Western blots were also performed using AS 1216. AS CI was used in Western blots to check for virus replication in *N. benthamiana*.

Reciprocal serological tests were done to test cross reactivity of WLMV, PRSV-W, ZYMV, WMV-2 and MWMV using SDS immunodiffusion and indirect ELISA. In some tests a comovirus, squash mosaic virus (SqMV), was used. AS 1216 from early and late bleedings was tested in Western blots against WLMV using the expressed coat

protein, purified virus and extracts of plants infected with WLMV, PRSV-W, PRSV-P, WMV-2, ZYMV and MWMV. Pre-immune serum and noninoculated or mock-inoculated plants were used as negative controls.

Reactivity of the polyclonal ascitic fluid from mouse number 2 (PAF #2) with WLMV was compared to the reactivity of AS 1216 #13 with WLMV in indirect ELISA and in Western blots. The expressed coat protein of WLMV and extracts of WLMV-infected squash were tested against both sera. Ascitic fluid from mouse number three, the nonimmunized mouse (PAF #3), and noninoculated or mock-inoculated EPSN squash were used as negative controls. PAF #2 was tested in indirect ELISA, TAS indirect ELISA, and in Western blots against WLMV, PRSV-W, WMV-2, ZYMV and MWMV. PAF #3 and noninoculated or mock inoculated EPSN were used as negative controls in indirect ELISA and in Western blots.

PTY-1 and ADMIX were tested against WLMV, PRSV-W, ZYMV, WMV-2 and MWMV in indirect ELISA and in Western blots. Indirect ELISA was also used to test the reactivity of PTY-1 with aphid-transmitted and mechanically-transmitted isolates of WLMV and PRSV-W.

Results

Expression of the Coat Protein Coding Region

Bacterial lysates collected 3 h after induction with IPTG gave clear precipitin lines when tested in immunodiffusion against AS 1163, whereas no precipitin lines were seen between AS 1163 and lysates of nontransformed or noninduced BL21 (DE3) pLysS. In Coomassie blue-stained SDS gels a clear band co-migrating with purified virus was seen

in samples of induced bacterial lysates; this band was not seen in samples of noninduced cultures (Fig. 4-1). In Western blots, a band which was clearly visible in the induced cultures, gave a strong positive reaction with AS 1163 (Fig. 4-2). This band was absent in noninduced cultures. Comparisons of the sequence of the recombinant plasmid used for expression of the coat protein with the sequences of clones pCPA and pCPB, showed that there were no PCR-induced mutations within the coat protein coding region of WLMV used for expression.

The antiserum prepared against the expressed coat protein (AS 1216) reacted against purified expressed coat protein, purified virus and extracts of plants infected with WLMV in immunodiffusion, Western blots, and in indirect ELISA. Side by side comparison of AS 1216 and an antiserum prepared to denatured purified virus (AS 1163) confirmed the identity of the expressed coat protein in Western blots (Fig. 4-3).

AS 1216 was titered by indirect ELISA to a dilution of 1/80,000. The titers of antisera were highest after the booster injection. OD_{405} values above 1.0 were recorded against purified expressed coat protein (100 ng/100 µl), and of 0.5 against infected leaf extracts diluted 1:10. OD_{405} values of 0.08 were recorded with extracts from healthy EPSN. In indirect ELISA, cross reactions were seen between WLMV and each of four potyviruses (PRSV-W, ZYMV, WMV-2 and MWMV), but not with the comovirus SqMV. Dilutions of AS 1216 within the range of 1/8,000 to 1/64,000 reduced cross reactivity of the antiserum and improved the specificity.

Host Range Tests

No symptoms were observed and no virus was recovered from the following plants when inoculated with WLMV: *Cassia occidentalis*, *Chenopodium amaranticolor*,

Chenopodium quinoa, *Phaseolus vulgaris* 'BT-1', 'BT-2' and 'Top crop', and *Nicotiana tabacum* 'Burley 21' and 'NN'. WLMV infections were verified by mechanical inoculation of EPSN indicator plants and also by indirect ELISA.

Symptoms in squash EPSN and 'Prelude II' were mosaic, and in 'Crimson Sweet' and 'Charleston Gray' watermelon, symptoms were vein clearing, mosaic, and leaf mottle. In Western blots of EPSN inoculated with WLMV, both coat protein and cylindrical inclusion protein were detected. When 'Prelude II' was tested in Western blots using two high titer collections of AS 1216, collection #13 (AS 1216 #13) and collection #15 (AS 1216 #15), a band of approximately 43 kDa in size was visible in the lanes loaded with extracts of healthy 'Prelude II' and 'Prelude II' infected with MWMV. Similar bands were seen when PTY-1, ADMIX, and PAF #2 were used. These bands were absent when healthy EPSN and EPSN infected with MWMV were used.

Cucumis metuliferus PI 482459 and PI 292190 reacted with chlorotic symptoms of the leaves and stems which later became necrotic. Localized necrosis was followed by death of the plant or recovery. Plants that recovered showed some veinal necrosis on leaves and necrotic streaks on the stems but were mostly symptomless. *Cucumis metuliferus* Acc 2459 showed mottle or mosaic. Although WLMV was recovered in EPSN from the three accessions of *C. metuliferus* tested, only Acc 2459 was consistently positive in indirect ELISA, and virus coat protein was not detected in Western blots (Tables 4-3 and 4-4).

Slight vein clearing was seen initially in the first leaves of *Lagenaria siceraria* but no symptoms were seen in leaves that developed later. WLMV, however, was detected in indirect ELISA. *Luffa aegyptiaca* showed rare yellow spots on leaves or was

symptomless, and WLMV was detected by indirect ELISA in the isolated yellow spots and in the green areas (Table 4-4). *Luffa acutangula* showed no symptoms. No virus was recovered from this host, or detected in indirect ELISA. *Cucurbita moschata* showed no symptoms and WLMV was not detected in indirect ELISA or recovered to squash in the spring test. However, although no symptoms were seen on the fall test, virus was detected in ELISA and recovered to squash (Table 4.4).

Momordica charantia and *Cucurbita okeechobeensis* var *okeechobeensis* showed no definite symptoms. Virus was recovered to EPSN squash, but was detected inconsistently in indirect ELISA. *Cucurbita ecuadorensis*, *Cucumis sativus* 'Surinam' and 'Taiwan TGM-1' showed no symptoms and WLMV was not detected in Western blots or in indirect ELISA. WLMV was not recovered in EPSN squash from *C. equadorensis* and *C. sativus* 'Surinam' (Table 4-3).

N. Benthamiana inoculated with WLMV showed no symptoms and although virus was recovered to EPSN squash from inoculated leaves, no virus was recovered from leaves above the inoculation point. No coat protein was detected in indirect ELISA. In Western blots of inoculated leaves and leaves above the point of inoculation, neither coat protein nor cylindrical inclusion protein were detected.

Serological Tests

In immunodiffusion tests of AS 1216 against WLMV in leaves of EPSN squash, watermelon and cantaloupe, clear precipitin lines were seen after 24 h with AS 1216 # 9. All subsequent antiserum collections reacted positively in immunodiffusion with WLMV in squash plants. In tests with antisera from earlier bleedings, precipitin lines were absent, or very faint after 48 h incubation. No reactions were seen against host plant proteins and

no cross reactions were seen in reciprocal immunodiffusion tests with WLMV, PRSV-W, ZYMV, WMV-2.

In indirect ELISA, WLMV cross-reacted with PRSV-W, ZYMV, WMV-2 and MWMV, but no cross reactions were seen between AS 1216 and SqMV. Dilutions of AS 1216 within the range of 1/8,000 to 1/64,000 reduced reactivity of the antiserum and improved the specificity of the test (Fig. 4-4).

In Western blots, AS 1216 from early and late bleedings recognized the coat proteins and possible breakdown products and dimers of the coat proteins of WLMV, PRSV-W, WMV-2, ZYMV, and MWMV. No similar proteins were detected in extracts from healthy EPSN tissues (Fig. 4-5) or from tissues infected with SqMV comovirus.

Mouse polyclonal antisera from test bleedings were titered by indirect ELISA against the expressed coat protein (5 µg/ml), and mouse number one (AS #1) was found to be more IgG active than mouse number two (AS #2). Although the titer against purified coat protein was higher, little ascitic fluid was collected from mouse number one making it necessary to conduct the tests with ascitic fluid from mouse number two (PAF #2).

In plate-bound antigen indirect ELISA and in Western blots, PAF #2 reacted with the expressed coat protein and with WLMV in infected plants, but not with noninfected or mock-inoculated plants. Absorbance values in ELISA after 60 min for PAF #2 against WLMV were much lower than those for AS 1216. A_{405} (PAF #2) = 0.221 and A_{405} (AS 1216) = 1.863 against expressed coat protein, and A_{405} (PAF #2) = 0.332 and A_{405} (AS 1216) = 1.807 against infected EPSN. Absorbance values found for PAF #2 against healthy EPSN were A_{405} = 0.012, and A_{405} = 0.342 for AS 1216. In Western blots, PAF

#2 reacted with WLMV in infected squash and faint bands were seen in lanes with ZYMV, MWMV and PRSV-W, but not with WMV-2 after incubation in the developing solution for more than 1 h (Fig. 4-6).

When PAF #2 was used to trap antigens in TAS indirect ELISA using AS 1216 #13 (1:1,000) as the secondary antibody, cross reactions with PRSV-W, WMV-2, ZYMV and MWMV and healthy EPSN were reduced considerably, making this the most specific assay procedure, second to immunodiffusion.

Reactions of PTY-1 with WLMV in indirect ELISA were inconsistent. A_{405} readings after 60 min varied from 0.033 to 0.448 depending on the sample tested (A_{405} values were more than twice the A_{405} values for noninoculated or mock-inoculated EPSN squash). A_{405} readings of samples infected with PRSV-W varied similarly. A_{405} readings for squash infected with ZYMV and WMV-2 were more consistent and were positive in all tests. When the aphid-transmitted culture and the mechanically-transmitted culture of WLMV were used, PTY-1 reacted with both cultures in two tests but not in another. The same observations were made when ADMIX was used, although readings were higher.

In Western blots using PTY-1, clear bands were seen against WMV-2 and ZYMV, and faint bands were seen against WLMV and PRSV-W after incubation in the developing solution for more than 1 h. In Western blots using ADMIX all bands were stronger, and WLMV and PRSV-W were detected by this test. No bands were seen in lanes containing samples from healthy or mock-inoculated EPSN squash (Fig. 4-7).

Discussion

Plants which give clear and consistent symptoms when inoculated with a virus are useful as indicator plants, particularly when a set of plants is available that reacts distinctly to different viruses. Consistent and clear symptoms resulting from inoculation with watermelon leaf mottle virus (WLMV) were found in 'Crimson Sweet' and 'Charleston Gray' watermelon, EPSN and 'Prelude II' squash and *Cucumis metuliferus* PI 292109, PI 482459 and Acc 2459. All other plants were either symptomless or showed inconsistent symptoms, which may have been dependent on environmental conditions. Symptoms in watermelon and squash were vein clearing followed by mosaic or mottle. *C. metuliferus* PI 292109, PI 482459 and Acc 2459 reacted with very characteristic symptoms, which may be useful in differential diagnostics. *Cucumis metuliferus* PI 482459 and PI 292190 reacted to inoculation with WLMV with chlorotic symptoms which became necrotic, leaves and stems became necrotic, and necrosis was followed by death of the plant or recovery. Plants that recovered showed some veinal necrosis on leaves and necrotic streaks on the stems but were mostly symptomless. *Cucumis metuliferus* Acc 2459 showed mild mottle or mosaic.

PI 292190 has not shown symptoms when inoculated with PRSV, it carries a single dominant gene (*Wmv 1*) which confers a high level of resistance to PRSV-W (Provvidenti and Robinson, 1974) and to PRSV-P (Provvidenti and Gonsalves, 1982). PI 292190 became infected with WMV-2 showing systemic mosaic and responding to inoculation with SqMV with chlorotic and local lesions (Provvidenti and Robinson, 1974).

Acc 2459 reacts to inoculation with PRSV-W with severe systemic symptoms (Provvidenti and Robinson, 1974) and to PRSV-P with symptoms varying from severe (PRSV-HA) to mild or moderate mottle (PRSV-Fla) depending on the isolate tested (Provvidenti and Gonsalves, 1982). Acc 2459 reacted with systemic mottle when infected with WMV-2. Chlorotic local lesions and systemic mottle were seen when plants were inoculated with CMV and SqMV.

Luffa acutangula did not show symptoms when inoculated with WLMV, and it does not show symptoms when inoculated with most strains of WMV-2, but shows mosaic when inoculated with PRSV-W. (De Sa *et al.*, 1989). Reaction to inoculation with ZYMV has been reported to vary according to the strain tested.

Nicotiana benthamiana was not infected systemically by WLMV, and in previous experiments it reacted to inoculation with WMV-2 with systemic mosaic and was symptomless when inoculated with PRSV-W (De Sa *et al.*, 1989). However *N. benthamiana* has been reported to be systemically infected by 200 different species and strains of potyviruses, including WMV-2 and ZYMV (Horvath, 1993).

C. metuliferus Acc 2459 and PI 292190, *Luffa acutangula*, and *N. benthamiana* may be used together to differentiate between PRSV-W and PRSV-P, WMV-2 and WLMV. Response of *L. acutangula* to inoculation with ZYMV varies from mosaic to no symptoms, depending on the strain used. The responses of Acc 2459 and PI 292190 have not been described for different strains of ZYMV. However, WLMV does not induce distinctive symptoms on several cultivars of melon or cucumber and *C. okeechobeensis* var *okechobeensis* and *Lagenaria siceraria* (Purcifull *et al.*, 1998), whereas ZYMV induces mosaic in all these hosts. These plants may be used to differentiate WLMV from

ZYMV, and from MWMV. Using commercially available seeds of melon and cucumber may be particularly useful for initial identification of viruses, because this bypasses seed importation and quarantine problems.

Antiserum 1216 was prepared to the denatured expressed coat protein of WLMV. Denatured immunogens generally result in the production of antisera reactive towards sequential determinants, which are usually best for immunoblotting of SDS-PAGE separated proteins (Dunbar and Schoebel, 1992). AS 1216 was cross-reactive with all the potyviruses tested in Western blots, particularly serum from later bleedings which had higher titers. Because coat protein size differs among the potyviruses tested, it was possible to differentiate the viruses based on coat protein size and the intensity of the band.

AS 1216 also cross-reacted in indirect ELISA with all the potyviruses tested but not with the comovirus SqMV. However, at dilutions of 1/8,000 to 1/64,000 and accompanied by positive controls, AS 1216 has the potential to be used as a specific tool for serological detection of WLMV in field surveys. Lyophilized expressed coat protein is particularly useful as a positive control for detection of the virus in areas where it has never been previously described, eliminating the need for introduction of infected plant material.

Polyclonal ascitic fluid (PAF #2) was a very specific probe for detecting WLMV in infected plants in Western blots, and although it recognized coat proteins of the other potyviruses tested, bands were very faint compared to WLMV coat protein. The specificity of indirect ELISA was much greater when antigens were trapped with PAF #2 in TAS indirect ELISA. In indirect ELISA, absorbance readings were very low when

compared to AS 1216, and PAF #2 is not recommended for this assay. Ascitic fluid is obtained in small amounts of 2 to 5 ml from each mouse, and although it may not be practical to produce ascitic fluid for large-scale projects, it may be extremely useful for studies of proteins which are difficult to express in *E. coli*. Very small amounts of protein are needed to immunize the mice, several mice can be inoculated with minimal amounts of protein, and ascitic fluid may be obtained within a short period of time.

Reactions of PTY-1 with WLMV were inconsistent in indirect ELISA, and readings varied from A_{405} = 0.033 to 0.448 depending on the sample tested, but were always above twice the absorbance reading of samples from healthy or mock-inoculated EPSN squash. The lowest readings may have been associated with lower virus titers in the plants, judging by absorbance readings of the same samples with the homologous antisera. Assays done with the same samples using ADMIX also showed similar variability. Isolate W1-A of PRSV-W, ZYMV and WMV-2 were used in these tests, and similar results were observed with PRSV-W, but not with ZYMV and WMV-2. In indirect ELISA, isolate W1-A of PRSV-W has been previously reported to react positively with PTY-1. Absorbance values obtained from tests using three isolates of PRSV-W that did not react with PTY-1 were the same as for healthy controls (Baker *et al.*, 1991). The putative epitope recognized by PTY-1 may be located in the heptapeptide WTMMDGE, as indicated by analysis of deletion mutants of bean yellow mosaic virus isolate GDD (BYMV-GDD), within the highly conserved core region of the coat protein (Jordan, 1992; Torrance, 1995). It has been suggested that one or more of the contact amino acid residues are different in virus isolates not recognized by PTY-1 (Jordan, 1992). In WLMV, this putative epitope has a sequence WTMMDGT. The difference in

the sequence and in virus titer in the plant may explain the differences in the absorbance readings. The original published sequence of isolate W1-A is WVMMDGE (Quemada *et al.*, 1990). However, the isolate used in the present study has been mechanically transferred for several generations and may have undergone mutations, which together with the difference in titer, may also explain the variability in ELISA absorbancy readings.

PTY-1 and ADMIX did not give reliable results in Indirect ELISA and in Western blots, and use of these monoclonal antisera is not recommended for detection of WLMV.

Table 4-1. Viruses in the *Potyvirus* genus infecting plants in the family Cucurbitaceae, showing name of the virus, place,date and author of first report.

Bryonia Mottle Virus [c]	Morocco	1979	Lockhart and Fischer
Bryony Potyvirus I [c]	Italy	1980	Milne et al.
Bryony Potyvirus II [c]	Italy	1980	Milne et al.
Bryony Potyvirus III [c]	Italy	1980	Milne et al.
Caserta Mosaic Virus [c]	Brazil	1997	Lima et al.
Clover Yellow Vein Virus [a]	Italy	1965	Lisa and Dellavalle
Cucurbita Mosaic Virus [c]	Ivory Coast	1980	Fauquet and Thouvenel [d]
Melon Vein-Banding Mosaic Potyirus [b]	Taiwan	1993	Huang et al.
Melothria Mottle Virus [c]	Japan	1993	Yonaha et al.
Morocco Watermelon Mosaic Virus [b]	Morocco	1974	Fischer and Lockhart
Papaya Ringspot Virus [a]	U.S.A.	1949	Jensen
Squash Yellow Leaf Curl Virus [c]	Oman	1998	Zouba et al.
Telfairia Mosaic Virus [a]	Nigeria	1975	Nwauzo and Brown
Trichosanthes Mottle Virus [b]	Japan	1988	Yonaha et al.
Watermelon Mosaic Virus 2 [a]	U.S.A.	1965	Webb and Scott
White Bryony Potyvirus [b]	Italy	1980	Milne et al.
Zucchini Yellow Fleck Virus [a]	Italy	1981	Vovlas et al.
Zucchini Yellow Mosaic Virus [a]	Italy	1981	Lisa et al.

[a] definite species within the genus.
[b] tentative species within the genus.
[c] not listed in the Sixth ICTV Report.
[d] *in* Lovisolo and Lisa, 1983.

Table 4-2. Virus isolates and polyclonal rabbit antisera used in SDS immunodiffusion, ELISA and Western blots, for serological characterization of watermelon leaf mottle virus (WLMV).

VIRUS ISOLATES [a]	ANTISERUM	COLLECTION DATE
ZYMV (FC 1119)	AS 1133	09-26-1986
WMV-2 (FC 1656)	AS 1134	03-10-1988
PRSV-W (W1-A)	AS 1142	09-14-1988
MWMV (MOR-ATCC PV957)	AS 955	02-08-1980
WLMV (FC 2932)	AS 1163	01-31-1991
WLMV (FC 2932)	AS 1216	05-21-1997
SqMV (FC 979)	nt	nt

nt = not tested

[a] Zucchini yellow mosaic virus (ZYMV), watermelon mosaic virus 2 (WMV-2), papaya ringspot virus type W PRSV-W, Moroccan watermelon mosaic virus (MWMV), squash mosaic virus (SqMV).

Table 4-3. Table showing detection of watermelon leaf mottle virus (WLMV), in cucurbit plants tested in host range tests by Western blots, indirect ELISA, and recovery on squash 'Early Prolific Straightneck' (EPSN). Western blots were performed using purified IgG of AS 1163, tests were conducted at least twice for all plants except for *C. sativus* 'Surinam' and 'Taiwan TGM-1'.

Host Plant	Western Blot [a]	Indirect ELISA [b] (A_{405})	Recovery on EPSN
C. lanatus 'Crimson Sweet'	++	1.693	yes
C. pepo EPSN	++	1.299	yes
C. pepo 'Prelude II'	++	0.841	yes
C. metuliferus Acc. 2459	-	0.063	yes
C. metuliferus PI 292190	-	0.019	yes
C. sativus 'Taiwan TGM-1'	-	0.003	yes
C. sativus 'Surinam'	-	0.001	no

[a] Intensity of the bands was determined by visual assessment
[b] Absorbance values were recorded one hour after addition of the substrate. Values represent average of two duplicate wells.
++ = bands corresponding to virus coat protein
- = no visible virus bands

Table 4-4. Plants tested as hosts of watermelon leaf mottle virus (WLMV) using plate-bound antigen indirect ELISA using purified IgG from AS 1163, diluted 1:1,000 and AS 1216 diluted 1:8,000. Composite leaf samples were extracted 1:10 in extraction buffer. Plants considered as hosts were those having ELISA readings greater than two times the A_{405} of the non-inoculated plants. Attempts to recover of WLMV from inoculated test plants were made using squash 'Early Prolific Straightneck 'EPSN' by mechanical inoculation. *okeechobeensis* var *okeechobeensis* PI 561112, *Lagenaria siceraria* GRIF 11938, *Luffa aegyptiaca* PI 555606, *Momordica charantia* PI 489700. All other seeds were obtained from the Plant Pathology Dept., University of Florida

nt = not tested.
host = plants with A_{405} above two times that of noninoculated plants.
nonhost = plants with A_{405} readings below two times that of noninoculated plants.
Plants obtained from the NCRPIS: *Cucumis metuliferus* PI 2922190, PI 482459, *Cucurbita ecuadorensis* GRIF 9446; *Cucurbita*.

Plant Tested	Spring 1997		Summer 1997			Fall 1997		
	IgG 1163	Recovery	IgG 1163	AS 1216	Recovery	IgG 1163	AS 1216	Recovery
C. lanatus 'Crimson Sweet'	host	yes	nt	nt	nt	nt	nt	nt
C. lanatus 'Charleston Gray'	nt	nt	nt	nt	nt	host	host	yes
Cucurbita pepo 'EPSN'	host	nt	host	host	nt	host	host	nt
Cucurbita pepo 'Prelude II'	host	yes	nt	nt	nt	host	host	yes
Cucumis metuliferus Acc 2459	host	yes	host	host	yes	host	host	yes
Cucumis metuliferus PI 292190	host	yes	nonhost	nonhost	yes	nonhost	nonhost	yes
Cucumis metuliferus PI 482459	nt	nt	host	host	yes	nonhost	nonhost	yes
Cucumis sativus 'Surinam'	nonhost	no	nt	nt	nt	nt	nt	nonhost
Cucumis sativus 'Taiwan TMG-1'	nonhost	yes	nt	nt	nt	nt	nt	nt
Cucurbita ecuadorensis	nonhost	no	nonhost	nonhost	no	nt	nt	nt
C. okeechobeensis var *okeechobeensis*	nt	nt	nonhost	host	yes	host	nonhost	yes
Cucurbita moschata	non host	no	nt	nt	nt	host	host	yes
Lagenaria siceraria	nt	nt	host	host	yes	host	host	yes
Luffa acutangula	nt	nt	nt	nt	nt	nonhost	nonhost	no
Luffa aegyptiaca	nt	nt	nt	host	yes	nt	nt	yes
Momordica charantia	nt	nt	host	nonhost	nt	nt	nt	nt

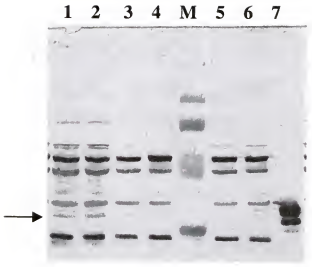

Fig. 4-1. SDS polyacrylamide gel (10%) followed by Coomassie blue staining for detection of the expressed coat protein of watermelon leaf mottle virus in lysates of *Escherichia coli* BL21 (DE3) pLys S. Lanes 1 and 2 = cultures expressing the coat protein after induction for three hours with IPTG. Lanes 3, 4, 5 and 6 = non-induced cultures. Lane 7 = purified virus preparation. Lane M contains the molecular weight markers (BRL) from top to bottom: Myosin (H-chain) 200 kDa, Phosphorylase B 97 kDa, Bovine serum albumin 68 kDa, Ovalbumin 43 kDa, and Carbonic anhydrase 29 kDa. Arrow indicates the expressed coat protein at ca 34 kDa.

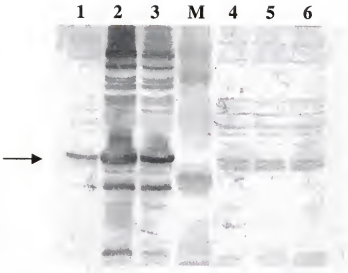

Fig. 4-2. Western blot of SDS polyacrylamide gel (10%) for detection of the expressed coat protein of watermelon leaf mottle virus (WLMV) in lysates of *E.coli* BL21 (DE3) pLys S. Membrane was probed with antiserum from denatured purified virions of WLMV (AS 1163). Lanes 1, 2 and 3 = cultures expressing the coat protein after induction for three hours with IPTG. Lanes 4, 5 and 6 = non-induced cultures. Lane M contains the molecular weight markers (BRL) from top to bottom: Phosphorylase B 97 kDa, Bovine serum albumin 68 kDa, Ovoalbumin 43 kDa, Carbonic anhydrase 29 kDa, and β - Lactoglobulin 18.4 kDa. Arrow indicates the expressed coat protein at ca 34 kDa.

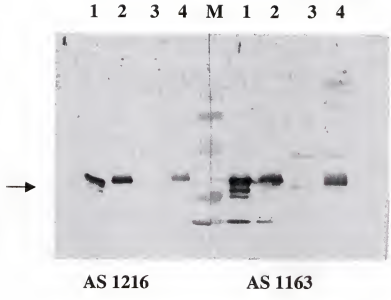

Fig. 4-3. Western blots of watermelon leaf mosaic virus (WLMV) coat protein probed with antiserum to the expressed coat protein (AS 1216) and antiserum to the dissociated purified virus (AS 1163). Proteins were transferred to a nitrocellulose membrane that was cut in two after the transfer, and each half was probed separately with the antisera. Lane 1 = purified virus; 2 = purified expressed coat protein; 3 = healthy 'Early Prolific Straightneck' (EPSN) squash; 4 = EPSN squash infected with WLMV. Protein standard markers (BRL) were from top to bottom in lane M: Myosin (H- chain) 200 kDa, Phosphorylase B 97 kDa, Bovine serum albumin 68 kDa, Ovalbumin 43 kDa, Carbonic anhydrase 29 kDa, β - Lactoglobulin 18.4 kDa. Arrow indicates coat protein position at ca 34 kDa.

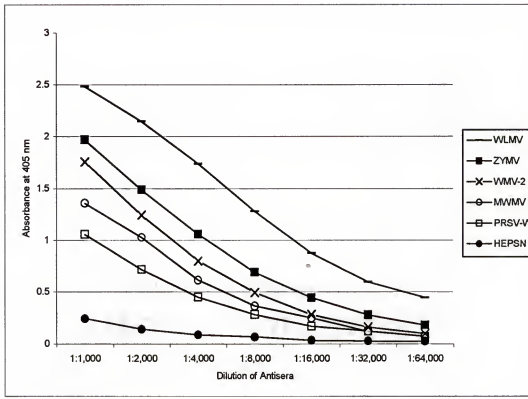

Fig. 4-4. Absorbance readings in plate-bound antigen indirect ELISA of rabbit polyclonal antiserum (AS 1216 coll. 05/21/97) prepared to the expressed coat protein of watermelon leaf mottle virus (WLMV) tested against WLMV, zucchini yellow mosaic virus (ZYMV), watermelon mosaic virus 2 (WMV-2), Moroccan watermelon mosaic virus (MWMV), papaya ringspot virus type W (PRSV-W), and healthy 'Early Prolific Straight Neck' (EPSN) squash.

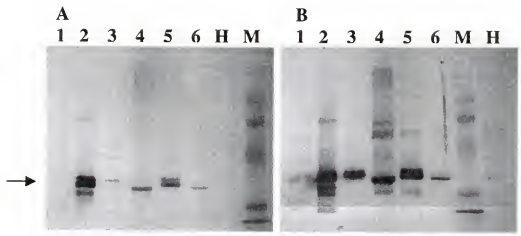

Fig. 4-5 Western blots of five potyviruses and one comovirus infecting 'Early Prolific Straightneck' (EPSN) squash probed with antiserum prepared to the expressed coat protein (AS 1216) of watermelon leaf mottle virus (WLMV). Membrane A was probed with antiserum collected five weeks after immunization (AS 1216 number 5), and membrane B was probed with antiserum collected six weeks after the booster injection (AS 1216 number 13). Tests were done simultaneously and antisera were diluted 1:2,000. Lanes 1 to 6 were loaded with extracts of EPSN squash infected with 1 = squash mosaic virus, 2 = WLMV, 3 = Morocco watermelon mosaic virus, 4 = zucchini yellow mosaic virus, 5 = papaya ringspot virus type W, 6 = watermelon mosaic virus 2.
Lane M = protein standard markers. Lane H = healthy EPSN.
Arrow indicates the coat protein of WLMV with an estimated molecular size of 34 kDa. A nonspecific reaction can be seen on lane 1 of Western blot B. Protein standard markers (BRL) were from top to bottom: Myosin (H-chain) 200 kDa, Phosphorylase B 97 kDa, Bovine serum albumin 68 kDa, Ovalbumin 43 kDa, Carbonic anhydrase 29 kDa, and β - Lactoglobulin 18.4 kDa.

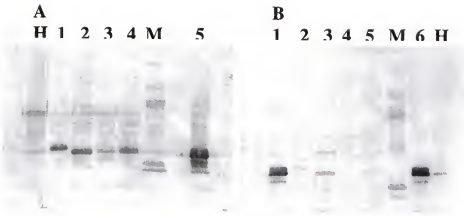

Fig. 4-6. Western blots of five potyviruses infecting 'Early Prolific Straightneck' (EPSN) squash, probed with rabbit polyclonal antiserum (AS 1216) and ascitic fluid (PAF #2) prepared to the expressed coat protein of watermelon leaf mottle virus (WLMV). Tests were done simultaneously. AS 1216 was diluted 1:2,000 and PAF #2 was diluted 1:1,000.
Membrane A was probed with AS 1216. Lanes H = healthy EPSN, 1 = Morocco watermelon mosaic virus, 2 = zucchini yellow mosaic virus, 3 = watermelon mosaic virus 2, 4 = papaya ringspot virus 2, 5 = WLMV, M = protein standards.
Membrane B was probed with PAF #2. Lanes 1 and 6 = WLMV, 2 = Morocco watermelon mosaic virus, 3 = zucchini yellow mosaic virus, 4 = watermelon mosaic virus 2, 5 = papaya ringspot virus type W, H = healthy EPSN (non specific reaction). M = protein standards.
Protein standard markers (BRL) were from top to bottom: Myosin (H-chain) 200 kDa, Phosphorylase B 97 kDa, Bovine serum albumin 68 kDa, Ovalbumin 43 kDa, Carbonic anhydrase 29 kDa, and β - Lactoglobulin 18.4 kDa.

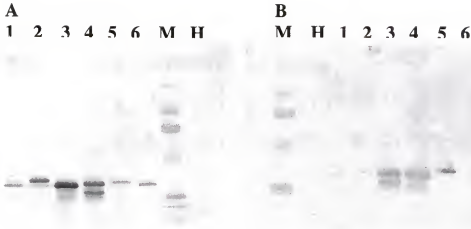

Fig. 4-7. Western blots of five potyviruses infecting 'Early Prolific Straightneck' (EPSN) squash, probed with an admixture of monoclonal antisera including PTY-1, PTY-2, PTY-4, PTY-8, PTY-10, PTY-21, PTY-24 and PTY-30 (ADMIX) and PTY-1. Tests were done simultaneously, and monoclonal antisera were diluted 1:1,000.
Membrane A was probed with ADMIX. Lanes 1 and 6 = watermelon leaf mottle virus, 2 = Morocco watermelon mosaic virus, 3 = zucchini yellow mosaic virus, 4 = watermelon mosaic virus 2, 5 = papaya ringspot virus type W, M = protein standards, H = healthy EPSN.
Membrane B was probed with PTY-1. Lanes M = protein standards, H = healthy EPSN, 1and 6 = watermelon leaf mottle virus, 2 = papaya ringspot virus type W, 3 = watermelon mosaic virus 2, 4 = zucchini yellow mosaic virus, 5 = Morocco watermelon mosaic virus.
Protein standard markers (BRL) were from top to bottom: Myosin (H-chain) 200 kDA, Phosphorylase B 97 kDa, Bovine serum albumin 68 kDa, Ovalbumin 43 kDa, Carbonic anhydrase 29 kDa, and β - Lactoglobulin 18.4 kDa.

CHAPTER 5
SUMMARY AND CONCLUSIONS

Biological, serological and molecular studies were carried out with watermelon leaf mottle virus (WLMV), a newly identified potyvirus infecting watermelon in Florida. Clones representing the entire genome of WLMV were isolated by generating two cDNA libraries in Lambda Zap II, by polymerase chain reaction (PCR) amplification of selected regions of the genome and by 5' rapid amplification of cDNA ends (5'RACE).

The nucleotide sequence derived from the cDNA clones indicated that the genome of WLMV is 9,994 nucleotides (nt) long. The single open reading frame (ORF) of 9,657 nt codes for a polyprotein of 3,218 amino acids (aa). The ORF is preceded by a 5' untranslated region (5'UTR) of 94 nt. The ORF ends in a UAA termination codon and is followed by a 3' untranslated region (3'UTR) of 243 nt. The sequence derived from the WLMV clones included 23 adenosine residues at the 3' terminus.

The taxonomic status of WLMV was determined by comparisons of nucleotide and amino acid sequences of the cloned viral genome with potyvirus sequences in GenBank. Similarity searches using the 3' terminal 2,316 nt of WLMV and the full-length polyprotein were performed using BLAST. Sequences of viruses with highest BLAST scores, of viruses infecting cucurbits, and of other potyviruses of economic importance were retrieved from GenBank. Multiple sequence alignments were done using PILEUP and CLUSTAL. Percentage of identity was calculated using GAP or ALIGN. Phylogenetic trees were generated using maximum parsimony and neighbor-

joining distance methods as implemented in PAUP and CLUSTAL, respectively. Optimal trees were subjected to bootstrap analysis as implemented in CLUSTAL with 1,000 replications, or as in PAUP with 100 replications.

Coat protein sequence identities between WLMV and 17 other members of the potyvirus genus ranging from 54.8 to 65.2% were found for amino acid sequences, and nucleotide sequence identities ranged from 55.6 to 63.5 %. Sequence identities between the 3'UTR of WLMV varied from 37.2 to 48.3%. The percentage of polyprotein amino acid sequence identity between WLMV and nine potyvirus sequences obtained from the polyprotein similarity search ranged from 46.4% to 48.3%. The percentage of sequence identity between WLMV and brome streak virus (BStV), a *Rymovirus*, was 27.9%. These values clearly indicate that WLMV is a unique species within the *Potyvirus* genus according to the criteria set forth by Shukla *et al.* (1988 and 1994) and Frenkel *et al.* (1989).

The topology of trees obtained from the coat protein and 3'UTR sequences, using neighbor-joining and maximum parsimony methods was similar as far as branches with high bootstrap support are concerned. Similar tree topology and the presence of branches already identified in the literature indicate the robustness of the trees. WLMV did not cluster with other viruses in branches with high bootstrap support and it was not possible to infer any phylogenetic relationships between WLMV and the other viruses.

The neighbor-joining tree generated from alignments of the polyprotein amino acid sequences of WLMV and nine other potyviruses, showed WLMV forming a clade with peanut stripe virus (PStV) and with the two strains of zucchini yellow mosaic virus (ZYMV). These viruses were also used in the phylogenetic analysis of the coat protein

and of the 3'UTR, but did not form a separate clade. The difference in tree topology of the polyprotein tree, and the coat protein and the 3'UTR tree, may be an indication of recombination events in the evolution of WLMV. The possibility of recombination is an area for further study of the evolution of WLMV as a species within the *Potyvirus* genus.

PSORT was used to predict cellular localization of viral proteins and to find possible nuclear localization signals (NLS). Previous studies by Petersen *et al.* (1996), using immunogold labeling with heterologous antisera prepared to nuclear and amorphous inclusions, have shown that conspicuous inclusions accumulate in the nucleus of squash and watermelon plants infected with WLMV.

PSORT prediction identified possible nuclear localization signals (NLS) in the P1 protein, helper component-protease (HC-Pro), the genome linked protein-protease (VPg-Pro), and in the replicase of WLMV. Possible NLS were also identified in the P1 proteins of PStV, plum pox virus (PPV), and tobacco etch virus (TEV). This is also an area of interest for further studies and may help identify the site of the potyviral P1 protein activity in the cell and its functions during the infection process. The P1 protein has not been localized to the nucleus of infected cells, but the HC-Pro of PPV, beet mosaic virus (BtMV) and PStV have been localized to the nucleus by immunogold labeling using antiserum produced to the expressed HC-Pro of PPV. Nuclear inclusions induced by WLMV were immunogold labeled with antiserum to the amorphous inclusions and may consist of the HC-Pro (Petersen *et al.*, 1996).

In nucleic acid-binding proteins, DNA or RNA binding motifs have been found to overlap with NLS within the same protein domains (LaCasse and Lefebvre, 1995). The presence of NLS in the P1 proteins of WLMV and of the other potyviruses mentioned

above, may be of interest due to the evidence that the P1 protein binds to nucleic acids (Soumounou and Laliberte, 1994). Additionally, there is evidence that the P1/HC-Pro sequence may function as a general suppressor of virus–induced gene silencing (Kasschau and Carrington, 1998), and as a broad range enhancer of pathogenicity (Pruss *et al.*, 1997).

The most efficient and widely used diagnostic procedures for virus identification are those based on the antigenic properties of the virus coat protein. In order to develop a serological probe for WLMV, the coat protein coding region of WLMV was PCR amplified and expressed in *Escherichia coli* BL12 (DE3) pLysS using the pETH expression vector. The denatured, purified expressed coat protein was used as an immunogen for the production of a polyclonal antiserum in rabbit and ascitic fluid in mice.

The antisera produced to WLMV were characterized using SDS-immunodiffusion, indirect ELISA, triple antibody sandwich (TAS) indirect ELISA, and Western blots for detection of WLMV in cucurbit plants. The commercially available monoclonal antiserum PTY-1, and an admixture of monoclonal antisera (ADMIX) were also tested to determine if they react with WLMV.

The polyclonal antiserum produced in a rabbit (AS 1216) was a specific probe for the detection of WLMV in immunodiffusion and may also be used for the detection of WLMV in ELISA and in Western blots. WLMV was detected in very low concentrations in inoculated plants using AS 1216 in Indirect ELISA. WLMV was found in immunodiffusion to be serologically distinct from papaya ringspot virus type W (PRSV-

W), watermelon mosaic virus (WMV-2), Moroccan watermelon mosaic virus (MWMV) and ZYMV.

The polyclonal ascitic fluid was useful for the detection of WLMV in Western blots and in indirect ELISA. Ascitic fluid was produced in less time and less expensively using 30 ug of purified protein. Only a small amount of the antiserum was produced and the titer of the antiserum produced was lower than the titer of AS 1216. In TAS indirect ELISA the specificity of indirect ELISA was increased when ascitic fluid was used to coat ELISA plates.

PTY-1 and ADMIX did not give reliable results in indirect ELISA and in Western blots, and use of these antisera is not recommended for detection of WLMV.

The availability of the cloned coat protein-coding region for production of test antigen bypasses the need to multiply the virus in plants. The bacterially-expressed coat protein may be used as a standard positive control for diagnosis in countries where the virus has not been described. This would avoid the introduction of infected plant material into an area where WLMV does not occur, and eliminate associated quarantine problems.

A host range study was performed to find plants useful as indicator hosts for diagnosis and differentiation of WLMV from other potyviruses infecting cucurbit plants.

Consistent and clear symptoms resulting from inoculation with WLMV were seen in 'Crimson Sweet' and 'Charleston Gray' watermelon, in 'Early Prolific Straightneck' (EPSN) and 'Prelude II' squash, and in *Cucumis metuliferus* PI 292109, PI 482459 and Acc 2459. *Luffa acutangula* and *Nicotiana benthamiana* did not show symptoms when inoculated with WLMV. *Cucurbita okeechobeensis* var *okeechobeensis*, *Lagenaria siceraria* and several cultivars of melon and cucumber, were either symptomless or

reacted with mild symptoms. These plants may be used for the initial identification of WLMV, particularly when control inoculations with the locally occurring potyviruses of cucurbits are performed.

A number of potyviruses that have not been sequenced yet, have been described infecting cucurbits in many parts of the world. Some of these viruses may be strains of the same virus and not new species. The sequences available for potyviruses infecting cucurbits are those of PRSV, several strains of ZYMV and partial sequences of WMV-2. The full-length sequence of WLMV adds to the database of potyviruses sequenced to date, and is the third potyvirus species infecting cucurbits to be entirely sequenced. The clones obtained can be ligated to make a full length infectious clone of WLMV, and the information presented here may enable experimentation that will elucidate some important aspects of viral pathogenesis.

LITERATURE CITED

Altschul, S.F., Gish, W., Miller, W., Myers, E.W., and Lipman, D.J. 1990. Basic local alignment search tool. J. Mol. Biol. 215: 430-10.

Anandalakashmi, R., Pruss, G.J., Ge, X., Marathe, R., and Mallory, A.C. 1998. A viral suppressor of gene silencing in plants. Proc. Natl. Acad. Sci. 95: 13079-13084.

Arbatova, J., Lehto, K., Pehu, E., and Pehu, T. 1998. Localization of the P1 protein of potato Y potyvirus in association with cytoplasmic inclusion bodies and in the cytoplasm of infected cells. J. Gen. Virol. 79:2319-2323.

Atreya, C.D., Atreya, P.L., Thornbury, D.W., and Pirone, T.P. 1992. Site-directed mutations in the potyvirus HC-Pro gene affect helper component activity, virus accumulation, and symptom expression in infected tobacco plants. Virology 191: 106-111.

Atreya, P.L., Lopez-Moya, J.J., Chu, M., Atreya, C.D., and Pirone, T. 1995. Mutational analysis of the coat protein N-terminal amino acids involved in potyvirus transmission by aphids. J. Gen. Virol. 76:265-270.

Baker, C.A. 1989. Production and characterization of polyclonal and monoclonal antibodies to three virus-induced proteins of papaya ringspot virus type W. Ph.D. Dissertation . University of Florida, Gainesville. 162p.

Baker, C.A., Lecoq, H., and Purcifull, D.E. 1991. Serological and biological variability among papaya ringspot virus type W isolates in Florida. Phytopathology 81:722-728.

Basso, J., Dallaire, P., Charest, P.J., Devantier, Y., and Laliberte, J.F. 1994. Evidence for an internal ribosomal entry site within the 5' non-translated region of turnip mosaic potyvirus RNA. J. Gen. Virol 75:3157-3165.

Berger, P.H., Wyatt, S.D., Shiel, P.J., Silbernagel, M.J., Druffel, K., and Mink, G.I. 1997. Phylogenetic analysis of the Potyviridae with emphasis on legume-infecting potyviruses. Arch. Virol. 142:1979-1999.

Blanc, S., Ammar, E.D., Garcia-Lampasona, S., Dolja, V. Llave, C., Baker, J., and Pirone, T.P. 1998. Mutation in the potyvirus helper component protein: effects on interactions with virions and aphid stylets. J. Gen. Virol. 79:3119-3122.

Brakke, M.K., and Van Pelt, N. 1970. Linear-log gradient for estimating sedimentation coefficients of plant viruses and nucleic acids. Anal. Biochem. 38:56-57.

Brantley, J.D. and Hunt, A.G. 1993. The N-terminal protein of the polyprotein encoded by the potyvirus tobacco vein mottling virus is an RNA-binding protein. J. Gen. Virol. 74: 1157-1162.

Carrington, J.C. and Freed, D.D. 1990. Cap-independent enhancement of translation by a plant potyvirus 5' nontranslated region. J. Virol. 64:1590-1597.

Carrington, J.C. and Herndon, K.L. 1992. Characterization of the potyviral HC-Pro autoproteolytic cleavage site. Virology 187:308-315.

Carrington, J.C., Freed, D.D., and Leinicke, A.J. 1991. Bipartite signal sequence mediates nuclear translocation of the plant potyviral NIa protein. The Plant Cell 3:953-962.

Chen, S., Das, P., Abdel Ghaffar, M.H., and Hari, V. 1995. Electron microscopic localization of ATPase activity in tobacco cells infected by tobacco etch potyvirus and tobacco mosaic virus. Arch. Virol. 140:173-178.

Christie, R.G. and Edwardson, J.R. 1994. Light and electron microscopy of plant virus inclusions. Florida Agricultural Experimental Stations Monograph No. 9 (revised). University of Florida, Gainesville.

Clark, M.F., and Adams, A.N. 1977. Characteristics of the microplate method of enzyme-linked immunosorbent assay for the detection of plant viruses. J. Gen. Virol. 34:475-483.

Converse, R.H., and Martin, R.R. 1990. ELISA methods for plant viruses. In: Hampton, R., Ball, E., and De Boer, S. (eds.) Serological Methods for Detection and Identification of Viral and Bacterial Plant Pathogens. A Laboratory Manual. APS. St. Paul, pp 179-204.

De Sa, P.B., and Hiebert, E. 1994. Cloning and characterization of the P3 protein of papaya ringspot virus type W (PRSV-W). (Unpublished manuscript).

De Sa, P.B., Hiebert, E., and Purcifull, D.E. 1996. Sequence analysis of the 3' terminus of a potyvirus isolated from watermelon in Florida. Phytopathology (Abstr.) 86:101.

De Sa, P.B., Dusi, A.N., Marinho, V.L.A., and Kitajima, E.W. 1989. Comparacao entre isolados do virus do mosaico do mamoeiro estirpe melancia (PRSV-W) e virus do mosaico da melancia 2 (WMV-2). Fitopatol. bras. (Abstr.) 14:144.

Dolja, V.V., Boyko, V. P., Agranovsky, A., and Koonin, E.V. 1991. Phylogeny of capsid proteins of rod-shaped and filamentous RNA plant viruses: two families with distinct patterns of sequence and probably structure conservation. Virology. 184: 79-86.

Dolja, V.V., Haldeman-Cahill, R., Montgomery, A.E., Vandenbosch, K.A., and Carrington, J.C. 1995. Capsid protein determinants involved in cell to cell and long distance movement of tobacco etch potyvirus. Virology 206:1007-1016.

Domier, L.L., Shaw, J.G., and Rhoads, R.E. 1987. Potyviral proteins share amino acid sequence homology with picorna-, como-, and caulimoviral proteins. Virology 158: 20-27.

Dunbar, B.S. and Schwoebel, E.D. 1992. Preparation of polyclonal antibodies. In: Deutscher, M.P. (ed.) Guide to Protein Purification. Methods in Enzymology 182:663-670.

Eagles, R.M., Balmori-Melian, E., Beck, D.L., Gardner, R.C., and Forster R.L.S. 1994. Characterization of NTPase, RNA-binding and RNA-helicase activities of the cytoplasmic inclusion protein of Tamarillo mosaic potyvirus. Eur. J. Biochem. 224:677-684.

Edwardson, J.R. 1992. Inclusion bodies. Arch. Virol. [Suppl. 5]:25-30.

Fernandez, A. and Garcia, J.A. 1996. The RNA helicase CI from plum pox potyvirus has two regions involved in binding to RNA. FEBS Letters 388:206-210.

Fischer, H.U. and Lockhart, B.E.L. 1974. Serious losses in cucurbits caused by watermelon mosic virus in Morocco. Pl. Dis. Reptr 58:143-146.

Frenkel, M.J., Ward, C.W., and Shukla, D.D. 1989. The use of 3' non coding nucleotide sequences in the taxonomy of potyviruses: application to watermelon mosaic virus 2 and soybean mosaic virus-N. J. Gen. Virol. 70:2775-2783.

Gallie, D.R. 1991. The cap and poly (A) tail function synergistically to regulate translation efficiency. Genes Dev. 5:2108-2116.

Gallie, D.R. Tanguay, R.L., and Leathers, V. 1995. The tobacco etch viral 5' leader and poly (A) tail are functionally synergistic regulators of translation. Gene 165:233-238.

Garcia, J.A., Martin, M.T., Cervera, M.T., and Riechmann, J.L. 1992. Proteolytic processing of the plum pox virus polyprotein at a novel cleavage site. Virology 188: 697-703.

Gotz, R. and Maiss, E. 1995. The complete nucleotide sequence and genome organization of the mite-transmitted brome streak mosaic rymovirus in comparison with those of potyviruses. J. Gen. Virol. 76:2035-2042.

Hajimorad, M.R., Ding, X.S., Flasinski, S., Mahajan, S., Graff, E., Haldeman-Cahill, R., and Carrington, J.C. 1996. NIa and NIb of peanut stripe potyvirus are present in the nucleus of infected cells, but do not form inclusions. Virology. 224:368-379.

Haldeman-Cahill, R., Daros, J.A., and Carrington, J.C. 1998. Secondary structures in the capsid protein coding sequence and the 3' nontranslated region involved in amplification of the tobacco etch virus genome. J. Virol. 72:4072-4079.

Hammond, J. 1992. Potyvirus serology, sequences and biology. Arch. Virol. [Suppl 5]: 123-138.

Hicks, G.R. and Raikhel, N.V. 1995. Nuclear localization signals binding proteins in higher plant nuclei. Proc. Natl. Acad. Sci. USA. 92: 734-738.

Hong, Y. and Hunt, A.G. 1996. RNA polymerase activity catalysed by a potyvirus-encoded RNA-dependent RNA polymerase. Virology 226:146-151.

Horvath, J. 1993. The role of *Nicotiana* species in plant virology with special regard to *Nicotiana benthamiana* Domin: a review. Acta Phytopath. et Entom. Hung. 28: 355-377.

Huang, C.H., Chang, L., and Tsai, J.H. 1993. The partial characterization of melon vein-banding mosaic virus, a newly recognized virus infecting cucurbits in Taiwan. Plant Pathology 42:100-107.

Huet, H., Gal-On, A., Meir, E., Lecoq, H., and Raccah, B. 1994. Mutations in the helper component protease gene of zucchini yellow mosaic virus affect its ability to mediate aphid transmissibility. J. Gen. Virol. 75:1407-1414.

Huynh, T.V., Young, R.A., and Davis, R.W. 1984. Constructing and screening cDNA libraries in lambda gt10 and lambda gt11. In: Glover, D. (ed.) DNA Cloning Techniques: a Practical Approach. IRL Press, Oxford. pp 49-78.

Ishihama, A., and Barbier, P. 1994. Molecular anatomy of viral RNA-directed RNA polymerases. Arch. Virol. 134:235-258.

Jensen, D.D. 1949. Papaya virus diseases with special reference to papaya ringspot. Phytopathology 39:191-211.

Jordan, R. 1992. Potyviruses, monoclonal antibodies, and antigenic sites. Arch. Virol. Suppl 5:81-95.

Jordan, R.L., and Hammond, J. 1991. Comparison and differentiation of potyvirus isolates and identification of strain-, virus-, subgroup-specific and potyvirus group-common epitopes using monoclonal antibodies. J. Gen Virol. 72: 25-36.

Kasschau, K.D., and Carrington, J.C. 1998. A counterdefensive strategy of plant viruses: suppression of postranscriptional gene silencing. Cell 95: 461-470.

Kasschau, K.D., Cronin, S., and Carrington, J.C. 1997. Genome amplification and long distance movement functions associated with the central domain of tobacco etch potyvirus helper component-proteinase. Virology 228:251-262.

Keller, K.E., Johansen, I.E., Martin, R.R., and Hampton, R.O. 1998. Potyvirus genome-linked protein (VPg) determines pea seed-borne mosaic virus pathotype-specific virulence in *Pisum sativum*. Mol. Plant-Microbe Interact. 11:124-130.

Kucharek, T., and Purcifull, D.E. 1997. Circular 1184. Florida Cooperative Extension Service, IFAS. University of Florida.

LaCasse, E.C., and Lefebvre, Y.A. 1995. Nuclear localization signals overlap DNA- or RNA-binding domains in nucleic acid-binding proteins. Nuc. Ac. Res. 10:1647-1656.

Laemmli, U.K. 1970. Cleavage of the structural proteins during the assembly of the head of the bacteriophage T4. Nature 227:680-685.

Lain, S., Martin, M.T., Riechman, J.L., and Garcia, J.A. 1991. Novel catalytic activity associated with positive-strand RNA virus infection: nucleic acid-stimulated ATPase activity of the plum pox potyvirus helicaselike protein. J. Virol. 65:1-6.

Langenberg, W.G. and Zhang, L. 1997. Immunocytology shows the presence of tobacco etch virus P3 protein in nuclear inclusions. J. Struct. Biol. 118:243-247.

Lecoq, H., and Pitrat, M. 1985. Specificity of the helper-component-mediated aphid transmission of three potyviruses infecting muskmelon. Phytopathology 75:890-893.

Lecoq, H., Wisler, G., and Pitrat, M. 1998. Cucurbit viruses: the classic and the emerging. In: J.D. Mc Creight (ed.). Cucurbitaceae 98: evaluation and enhancement of cucurbit germplasm. ASHS Press. Alexandria, VA.

Li, X.H., and Carrington, J.C. 1993. Nuclear transport of tobacco etch potyviral RNA-dependent RNA polymerase is highly sensitive to sequence alterations. Virology 103:951-958.

Li, X.H, Valdez, P., Olvera, R.E., and Carrington, J.C. 1997. Functions of the tobacco etch virus RNA polymerase (NIb): subcellular transport and protein-protein interaction with VPg/proteinase (NIa). J. Virol. 1598-1607.

Lima, J.A.A., do Vale, C.C., and de Oliveira, V.B. 1997. Viruses that infect cucurbits in the northeast of Brazil. Virus: Reviews and Research. J. Braz. Soc. Virol. 2:128-130.

Lisa, V., and Dellavalle, G. 1981. Characterization of two potyviruses in *Cucurbita pepo*. Phytopath. Z. 100:279-286.

Lisa, V., Boccardo, G., d'Agostino, G., Dellavalle, G., and d'Aquilio, H. 1981. Characterization of a potyvirus that causes zucchini yellow mosaic. Phytopathology 71:667-672.

Lockhart, B.E.L., and Fischer, H.U. 1979. Host range and some properties of *Bryonia* mottle virus, a new member of the potyvirus group. Phytopathol. Z. 96:244-250.

Lopez-Moya, J.J., and Pirone, T.P. 1998. Charge changes near the N terminus of the coat protein of two potyviruses affect virus movement. J. Gen. Virol. 79:161-165.

Lovisolo, O., and Lisa, V. 1983. Virosi e micoplasmosi delle Cucurbitacee. Estrato da L'Italia Agricola anno 120 no.1, gennaio-marzo.

Lutcke, H.A., Chow, K.C., Mickel, F.S., Moss, K.A., Kern, H.F., and Scheele, G.A. 1987. Selection of AUG initiation codon differs in plants and animals. EMBO J. 6: 43-48.

Maia, I.G., Haenni, A., and Bernardi, F. 1996. Potyviral HC-Pro: a multifunctional protein. J. Gen. Virol. 77:1335-1341.

Martin, M.T., and Gelie, B. 1997. Non-structural plum pox potyvirus proteins detected by immunogold labelling. Eur. J. Pl. Pathol. 103:427-431.

Mathews, R.E.F., 1992. Plant Virology. Fundamentals of Plant Virology. Academic Press, Inc., San Diego, CA. 405p.

McCarty, D.R., Hattori, T., Carson, C.B., Vasil, V., Lazar, M., and Vasil, I.K. 1991. The viviparous-1 developmental gene of maize encodes a novel transcriptional activator. Cell 66: 895-905.

Milne, R.G., Masenga, V., and Lovisolo, O. 1980. Viruses associated with white bryony (*Bryonia cretica* L.) mosaic in Northeastern Italy. Phytopathol. Medit. 19:115-120.

Murphy, F.A., Fauquet, C.M., Bishop, D.H.L., Ghabrial, S.A., Jarvis, A.W., Martelli, G.P., Mayo, M.A., Summers, M.D. (eds.). 1995. Virus Taxonomy. Sixth Report of the International Committee on Taxonomy of viruses. Springer-Verlag. Wien.

Murphy, J.F., Klein, P.G., R.E., Hunt., A.G., and Shaw, J.G. 1996. Replacement of the tyrosine residue that links the potyviral VPg to the viral RNA is lethal. Virology 220:535-538.

Murphy, J.F., Rychlik, W., Rhoads, R.E., Hunt. A.G. and Shaw, J.G. 1991. A tyrosine residue in the small nuclear inclusion protein of tobacco vein mottling virus links the VPg to the viral RNA. J. Virol. 65:511-513.

Nicolaisen, M., Johansen, E., Poulsen, G.B., and Borkhardt, B. 1992. The 5' untranslated region from pea seed-borne mosaic potyvirus RNA as a translational enhancer in pea and tobacco protoplasts. FEBS. 303:169-172.

Nwauzo, E.E. and Brown, W.M. 1975. Telfairia (Cucurbitaceae) mosaic virus in Nigeria. Pl. Dis. Reptr. 50:430-432.

Page, R. 1996. TreeView: an application to display phylogenetic trees on personal computers. Comput. Applic. Biosci. 12:357-358.

Peng, C.H., Kadoury, D., Gal-On, A., Wang, Y., and Raccah, B. 1998. Mutations in the HC-Pro gene of zucchini yellow mosaic potyvirus: effects on aphid transmission and binding to purified virions. J. Gen Virol. 79:897-904.

Petersen, M.A., Hiebert, E., and Purcifull, D.E. 1996. Immunolabeling of inclusion bodies induced by a potyvirus isolated from watermelon in Florida. (Abstr.) Phytopathology 86:99.

Provvidenti, R., and Gonsalves, D. 1982. Resistance to papaya ringspot virus in *Cucumis metuliferus* and its relationship to resistance to watermelon mosaic virus 1. J. Heredity 73:239-240.

Provvidenti, R., and Robinson, R.W. 1974. Resistance to squash mosaic virus and watermelon mosaic virus 1 in *Cucumis metuliferus*. Pl. Dis. Reptr. 58:735-738.

Pruss, G., Ge, X., Shi, X.M., Carrington, J.C., and Vance V.B. 1997. Plant viral synergism: the potyviral genome encodes a broad-range pathogenicity enhancer that transactivates the replication of heterologous viruses. The Plant Cell 9:859-868.

Purcifull, D.E. 1990. Ouchterlony double diffusion in the presence of sodium dodecyl sulfate for the detection of virion proteins and virus induced inclusion body proteins. In Hampton, R., Ball, E., De Boer, S. (eds). Serological Methods for Detection and Identification of Viral and Bacterial Plant Pathogens. A Laboratory Manual. APS. St. Paul, pp 121-127.

Purcifull, D.E. and Hiebert, E. 1979. Serological distinction of watermelon mosaic virus isolates. Phytopathology 69:112-116.

Purcifull, D.E., Adlerz, W.C., Simone, G.W., Hiebert, E., and Christie, S.R. 1984. Serological relationships and partial characterization of zucchini yellow mosaic virus isolates from squash in Florida. Plant Dis. 68:230-233.

Purcifull, D.E., Hiebert, E., Petersen, M.A., Simone, G.W., Kucharek, T.A., Gooch, M.D., Crawford, W.E., Beckham, K.A., and De Sa, P.B. 1998. Partial characterization of a potyvirus isolated from watermelon in Florida. Plant Dis. 82:1386-1390.

Quemada, H., Sieu, L. C., Siemieniak, D.R., and Gonsalves, D. 1990. Watermelon mosaic virus 2 and zucchini yellow mosaic virus: cloning of 3'terminal regions, nucleotide sequences, and phylogenetic comparisons. J. Gen. Virol. 71:1451-1460.

Restreppo, M.A., Freed, D.D., and Carrington, J.C. 1990. Nuclear transport of plant potyviral proteins. Plant Cell 2:987-998.

Riechmann, J.L., Lain, S., and Garcia, J.A. 1992. Highlights and prospects of potyvirus molecular biology. J. Gen. Virol. 73:1-16.

Riedel, D., Leseman, D.E., and Maiss, E. 1998. Ultrastructural localization of nonstructural and coat proteins of 19 potyviruses using antisera to bacterially expressed proteins of plum pox virus. Arch. Virol. 143:2133-2158.

Roberts, I.M., Wang, D., Findlay, K., and Maulet, A.J. 1998. Ultrastructural and temporal observations of the potyvirus cylindrical inclusion (CI) show that the CI protein acts transiently in aiding virus movement. Virology 245:173-181.

Rodriguez-Cerezo, E. and Shaw, J.G. 1991. Two newly detected nonstructural viral proteins in potyvirus-infected cells. Virology 185: 572-579.

Rodriguez-Cerezo, E., Gamble Klein, P., and Shaw, J.G. 1991. A determinant of disease symptom severity is located in the 3' terminal noncoding region of the RNA of a plant virus. Proc. Nat. Acad. Sci. USA. 88:9863-9867.

Rodriguez-Cerezo, E., Ammar, E.D., Pirone, T.P., and Shaw, J.G. 1993. Association of the non-structural P3 viral protein with cylindrical inclusions in potyvirus-infected cells. J. Gen. Virol. 74:1945-1949.

Rost, B., and Valencia, A. 1996. Pitfalls of protein sequence analysis. Current Opinions in Biotechnology 7:457-461.

Sambrook J., Fritsch, E.F., and Maniatis, T. 1989. Molecular Cloning : A Laboratory Manual. 2nd ed., Vol. 2. Cold Spring Harbor Laboratory, Cold Spring Harbor, NY.

Sanger, F., Nicklen, S., and Coulson, A.R. 1977. DNA sequencing with chain terminating inhibitors. Proc. Natl. Acad. Sci. USA 74: 5463-5467.

Schaad, M.C., Haldeman-Cahil, R., Cronin, S., and Carrington, J.C. 1996. Analysis of the VPg-proteinase encoded by tobacco etch potyvirus: effects of mutations on subcellular transport, proteolytic processing and genome amplification. J. Virol. 70:7039-7048.

Schaad, M.C., Jense, P.E., and Carrington, J.C. 1997a. Formation of plant RNA virus replication complexes on membranes: role of an endoplasmic reticulum-targetted viral protein. Embo J. 16:4049-4059.

Schaad, M.C., Lellis, A.D., and Carrington, J.C. 1997b. VPg of tobacco etch potyvirus is a host genotype specific determinant for long-distance movement. J. Virol. 71:8624-8631.

Schaeffer, B.C. 1995. Revolutions in rapid amplification of cDNA ends: new strategies for polymerase chain reaction cloning of full-length cDNA ends. Anal. Biochem. 227:255-273.

Shi, X.M., Miller, H., Verchot, J., Carrington, J.C., and Vance, V.B. 1997. Mutations in the region encoding the central domain of helper component-proteinase (HC-Pro) eliminate potato virus X/potyviral synergism. Virology 231:35-42.

Shukla, B.D. and Ward, C.W. 1988. Amino acid sequence homology of coat proteins as a basis for identification and classification of the potyvirus group. J. Gen. Virol. 69:2703-2710.

Shukla, D.D., Ward, C.W., and Brunt, A.A. 1994. Potyviridae. CAB International, Wallingford, UK. 516 pp.

Simon-Buela, L., Guo, H.S., and Garcia, J.A. 1997a. Cap-independent leaky scanning as the mechanism of translation initiation of a plant viral genomic RNA. J. Gen. Virol. 78:2691-2699.

Simon-Buela, L., Guo, H.S., and Garcia, J.A. 1997b. Long sequences in the 5' noncoding region of plum pox virus are not necessary for viral infectivity but contribute to viral competitiveness and pathogenesis. Virology 233:157-162.

Soumounou, Y., and Laliberte, J.F. 1994. Nucleic acid-binding properties of the P1 protein of turnip mosaic virus potyvirus produced in *Escherichia coli*. J. Gen. Virol. 75:2567-2573.

Studier, F.W., Rosenberg, A.H., Dunn, J.J., and Dubendorff, J.W. 1990. Use of T7 RNA polymerase to direct expression of cloned genes. Methods in Enzymology 185: 60-88.

Swoford, D.L. 1993. PAUP: phylogenetic analysis using parsimony, version 3.1.1. Illinois Natural History Survey, Champaign. Illinois.

Tanguay, R.L., and Gallie, D.R. 1996. The effect of the length of the 3' untranslatable region on expression in plants. FEBS Letters 394:285-288.

Thompson, J.D., Higgins, D.G., and Gibson, T.J. 1994. CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. Nucl. Acids Res. 22:4673-4680.

Thornbury, D.W., Hellmann, G.M., Rhoads, R.E., and Pirone, T.P. 1985. Purification and characterization of potyvirus helper component. Virology 144:260-267.

Torrance, L. 1995. Use of monoclonal antibodies in plant pathology. Eur. J. Plant Pathology 101:352-363.

Towbin, H. T., Staelin, T., and Gordon, J. 1979. Electrophoretic transfer of proteins from polyacrylamide gels to nitrocellulose sheets: procedure and some applications. Proc. Nat. Acad. Sci. USA 76: 4350-4354.

Van Regenmortel, M.H.V. 1982. Serology and Immunochemistry of Plant Viruses. Academic Press, Inc., New York. 302pp.

Verchot, J., and Carrington, J.C. 1995. Evidence that potyvirus P1 protein functions in *trans* as an accessory factor for genome amplification. J. Virol. 69:3668-3674.

Verchot, J., Koonin, E.V., and Carrington, J.C. 1991. The 35 kDa protein from the N-terminus of the potyviral polyprotein functions as a third virus encoded proteinase. Virology 185:527-535.

Vovlas, C. Hiebert, E., and Russo, M. 1981. Zucchini yellow fleck virus, a new potyvirus of zucchini squash. Phytopathol. Medit. 20:123-128.

Webb, R.E., and Scott, H.A. 1965. Isolation and identification of watermelon mosaic viruses 1 and 2. Phytopathology 55:895-900.

Wisler, G.C. 1992. Characterization of the P1 protein of the zucchini yellow mosaic potyvirus. Ph.D. Dissertation . University of Florida, Gainesville. 129p.

Xiao, X.W., Frenkel, M.J., Teakle, D.S., Ward, C.W., and Shukla, D.D. 1993. Sequence diversity in the surface exposed amino-terminal region of the coat protein of seven strains of sugarcane mosaic virus correlates with their host range. Arch. Virol. 132: 399-408.

Yonaha, T., Tamori, M., Fujiwara, Y., and Ishimine, M. 1988. Studies on potyvirus parasite on Cucurbitaceae in Okinawa. 1. Some properties of Trichosanthes mottle virus and watermelon mosaic virus 1. Bull. Coll. Agric. Ryukyus 35:1-15.

Yonaha, T., Nagata, M., Kawano, S., and Tamori, M. 1993. Studies on potyvirus parasite on Cucurbitaceae in Okinawa. 2. Some properties of Melothria mottle virus. Bull. Coll. Agric. Ryukyus 40:9-19.

Zouba, A.A. and Lopez, M.V. 1998. Squash yellow leaf curl virus: a new whitefly-transmitted poty-like virus. Plant Dis. 82:475-478.

Zuker, M. 1989. On finding all suboptimal foldings of an RNA molecule. Science 244: 48-52.

BIOGRAPHICAL SKETCH

Patricia de Sa Guimaraes (nee Boggan de Sa) was born in Recife, Pernambuco, Brazil. She graduated from the Federal University of Pernambuco, Recife, Brazil, with a Bachelor of Science degree in biological sciences in 1975. Patricia spent the next years caring for her family, and was involved with the needs and cultures of the indigenous peoples of Brazil.

She returned to the pursuit of her interest in plant sciences at the University of Brasilia, with a scholarship from 'CNPq-Conselho Nacional de Pesquisa e Pos Graduacao', Brazil. She graduated from the University of Brasilia in 1989 with a Master of Science degree in plant pathology. She worked for three years at the University of Brasilia at the Electron Microscopy Laboratory of the Department of Cellular Biology, and at the Department of Plant Pathology. She was sponsored by CNPq with a Regional Scientific Development Grant to carry out research on the characterization of plant viruses. Patricia was sponsored by the CNPq to begin a doctoral program at the University of Florida, under the direction of Dr. Ernest Hiebert and Dr. Dan E. Purcifull and expects to receive the degree of Doctor of Philosophy in plant pathology in May 1999.

Patricia was the Graduate Student Council Representative for the Department of Plant Pathology and a member of the IFAS Graduate Student Advisory Committee, at the University of Florida. She is a member of the Fraternity of Alpha Zeta, of Gamma Sigma Delta Honor Society, and a member of the American Phytopathological Society.

I certify that I have read this study and that in my opinion it conforms to acceptable standards of scholarly presentation and is fully adequate, in scope and quality, as a dissertation for the degree of Doctor of Philosophy.

Ernest Hiebert, Chairman
Professor of Plant Pathology

I certify that I have read this study and that in my opinion it conforms to acceptable standards of scholarly presentation and is fully adequate, in scope and quality, as a dissertation for the degree of Doctor of Philosophy.

Dan E. Purcifull, Cochairman
Professor of Plant Pathology

I certify that I have read this study and that in my opinion it conforms to acceptable standards of scholarly presentation and is fully adequate, in scope and quality, as a dissertation for the degree of Doctor of Philosophy.

Francis W. Zettler
Professor of Plant Pathology

I certify that I have read this study and that in my opinion it conforms to acceptable standards of scholarly presentation and is fully adequate, in scope and quality, as a dissertation for the degree of Doctor of Philosophy.

William B. Gurley
Associate Professor of Microbiology
and Cell Science

This dissertation was submitted to the Graduate Faculty of the College of Agriculture and to the Graduate School and was accepted as partial fullfillment of the requirements for the degree of Doctor of Philosophy.

May, 1999

Dean, College of Agriculture

Dean, Graduate School